# A Modified Viterbi Equalization Algorithm for Mitigating Timing Errors in Optical Turbulence Channels 

Yatian $\mathrm{Li}^{\ominus}$, Member, IEEE, Tianwen Geng ${ }^{\oplus}$, Ruotong Tian ${ }^{\ominus}$, and Shijie Gao


#### Abstract

In high-speed optical wireless communication (OWC) systems, timing errors can blur the edges of the received symbols, defined by the deviation of a signal's timing event from its intended occurrence. The timing errors introduce the inter-symbol interference (ISI) to the neighboring symbols. In order to mitigate the degradation, a maximum likelihood (ML) based Timing Error Viterbi Equalization (TEVE) algorithm is proposed, in which the probability density functions (PDF) of the channel conditions and the timing errors are priori information. The mathematical form of each branch metric (BM) is deduced by those statistics in different situations of neighboring signs of timing errors. The summations of BMs form the cumulative metrics (CM). By adopting the add-compare-select iterations, we are able to detect symbols with the largest CMs. The theoretical closed-form average bit error rates (BER) are also deduced and compared in the cases whether the TEVE algorithm is utilized. Experimental results indicate that our proposed method can reduce timing errors' impairment significantly with an acceptable complexity.


Index Terms-Optical communication, intersymbol interference, timing jitter, maximum likelihood detection.

## I. Introduction

LASER communication systems are rapidly developed due to the advantages of tremendous bandwidth, security, license-free operation [1]-[5]. Compared with free space coherent optical communication, the intensity modulation and direct detection (IM/DD) system has more simplicity to implement [6], [7]. The optical link was achieved between the lunar orbiter and its ground segment (LLGT) with the data rate of 622 Mbps [9]. The laser power was only 0.5 W , while the optical antenna was as small as 10 cm [10]. Future plans for the next generation earth relay optical communications are in the process of design and development, where the space to

Manuscript received May 26, 2020; revised October 4, 2020; accepted January 6, 2021. Date of publication January 20, 2021; date of current version June 10, 2021. This work was supported by the Research Project of Scientific Research Equipment of the Chinese Academy of Sciences. The associate editor coordinating the review of this article and approving it for publication was X. Cheng. (Corresponding author: Yatian Li.)
Yatian Li, Tianwen Geng, and Shijie Gao are with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun 130033, China (e-mail: yt_li@ciomp.ac.cn).
Ruotong Tian is with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun 130033, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China.
Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3051161.

Digital Object Identifier 10.1109/TWC.2021.3051161
ground links from optical relay satellites will have the ability of processing data with the rate of 100 Gbps [11], [12].
Despite the bright future, there are also some inevitable challenges hampering the system performance when transmitting through the atmosphere [13]-[15]. On account of the eddies with different sizes causing refraction and scattering, the turbulence channel may lead to intensity scintillation and phase perturbation in the receiving end [16]. There have already been several statistical models to characterize the fading events. Popular models include Log-Normal [17], GammaGamma [18], Malaga [19], Rician-Lognormal model [20]. Besides the turbulence, misalignments also bring additional fading on the receiving power [21]. The random items of the misalignments follow Gaussian distributions in both horizontal and vertical direction. Depending on the mean values and variances of two directions, several models are elaborated, such as Hoyt [22], Rician [23], Beckmann distribution [24]. As a result, this article assumes that the fading channel consists of both turbulence and pointing errors.

Owing to the high transmission rate of FSO (free space optics), the accuracy of the clock is an additional important issue during the receiver's clock data recovery (CDR) progress. A timing error is defined as the deviation of a signal's timing event from its intended (ideal) occurrence in time [25]. According to Ref. [26], the timing errors can be modeled as Gaussian variables. The timing errors shifting the decision points can introduce an inevitable the inter-symbol interference (ISI) for the neighboring symbols [27]. Ref. [28] has evaluated the impairments of timing errors causing inter-carrier interference in the orthogonal frequency division multiplexing (OFDM) system, where the timing errors were converted into the complex phases by Fourier transform. Ref. [29] reported on the timing jitter characterization of the superconducting single flux quantum (SFQ) coincidence circuit, which is an essential component of the superconducting coincidence photon counter. Our previous works elaborated the system performance [31], [32]. Ref. [31] analyzed the timing errors' influence on the symbol error rate (SER) for the pulse-position modulation (PPM) scheme in the atmospheric turbulence channel. In Ref. [32], the ergodic capacity and outage performance have been analyzed in the aggregate channel including turbulence, misalignments and timing errors. Due to the fact that only one pulse exists in every PPM symbol, there are always no powers allocated on its neighboring slots, which cannot produce interference to the pulse slot in the
systems with guard slots. However, almost every bit suffers from ISI in the on-off keying (OOK) scheme with timing errors. This article discovers that the ISI has different forms due to the corresponding signs of adjacent timing errors, which has not been considered in our previous works. Therefore, we analyze the bit error rate (BER) performance in the cases of different timing errors' signs in this article. Both the upper and lower bounds are also derived, as well as the asymptotic bound.

As stated above, the timing errors can introduce ISI harming the system. Therefore, it is a natural approach to use equalization method to mitigate the ISI [6]. The literature of equalization in optical communication systems are first designed for the fiber channels, where the signal suffers from non-linear noise (NLN) and polarization mode dispersion (PMD) [33]-[36]. Kinds of solutions have been proposed such as zero-forcing (ZF) equalization [33], minimum mean square error (MMSE) equalization [34], Turbo equalization [35], maximum likelihood sequence detection (MLSD) equalization [36]. Then some recent works about equalizers have been done to fight against the ISI in the OWC system. A decision feedback equalizer (DFE) was proposed by Ref. [37], where a digital signal to noise ratio (DSNR) penalty model was utilized. With the help of the least mean square (LMS) algorithm, the adaptive DFE method was confirmed to increase the system throughput and maintain constant low BER. Ref. [38] presented the first optical equalizer realized on Indium Phosphide material that was also monolithically integrated with a semiconductor optical amplifier. The device was able to mitigate both inter-symbol interference arising from narrowband optical filtering of 40 Gbps non-return-to-zero (NRZ) data and the residual chirp and electrical filtering in the transmitter and receiver. Ref. [39] elaborated a graph-based equalization with low complexity by applying sum-product (SP) algorithm, which combined a recursive systematic convolutional code and PPM. However, according to Forney's valuable work [40], MLSD can be considered as the optimal equalizer in terms of minimizing the BER. Ref. [41] proposed a suboptimal low-complexity detection rule, based on the generalized maximum-likelihood sequence estimation (MLSE), where the channel was modeled as Poisson detection model. The scheme allowed the detection of sequence lengths to be prohibitive for conventional MLSD, without using any channel knowledge. Ref. [42] validated that the MLSD significantly outperformed the linear minimum mean square error (LMMSE), where a linear time-invariant (LTI) Poisson channel modeling method was adopted, and the ISI was based on a probabilistic delay profile. Ref. [30] proposed four detection methods for the OOK free space optical links under synchronization errors with high mobility. Two kinds of sequence data detectors with a single wavelength were designed based on ML algorithms and generalized likelihood ratio test (GLRT) criteria. Based on the differential signaling structure, another two solutions with lower computational complexity were proposed in the case of two wavelengths.

Motivated by the joint statistics of Viterbi equalization [36] and MLSD signal detection [30], [41], this article proposes a Timing Error Viterbi Equalization (TEVE) algorithm to fight
against the ISI brought by timing errors. It needs to be mentioned that this article divides the ISI into two scenes, which are the "clean event" and the "bad event", respectively. In the former case, we only consider the two adjacent symbols as the ISI items, while more symbols are taken into consideration in the latter case. The TEVE algorithm is designed for the "clean events", but it can also provide some help for "bad events". The receiver is assumed to obtain the statistics of compound channel gain $h$ and the timing error $\xi$. In our TEVE algorithm, the instantaneous timing errors' signs $\left\{\Xi\left(\xi_{k}\right)\right\}$ are indispensable constituent elements in all states of the TEVE trellis. Then the expressions of branch metrics (BM) and cumulative metrics (CM) are both deduced by the those statistics in four different cases of $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)= \pm \pm$. During the detection progress, the signal data can be detected by reserving the path with largest CM and discarding others. The main contributions of this article are mainly summarized as follows.
i) In terms of different adjacent $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)$, we evaluate the approximate BER $P_{s}$ expressions on the IM/DD atmospheric optical communication with timing errors. The upper and lower bounds are deduced in the closed form, as well as the asymptotic bound.
ii) In order to fight against the ISI caused by timing errors, the TEVE algorithm is formulated, which is based on the MLSD. Compared with traditional Viterbi Equalization (VE), the number of states has doubled, due to the positive and negative signs of timing errors. Besides both BMs and CMs should also consider the statistics of neighboring timing errors, as well as the statistics of channel gain and noise.
iii) The theoretical BER expression $P_{s}^{T E V E}$ after TEVE algorithm has been derived in a closed form, which are quite consistent with indoor experimental results.

For brevity, the remainder of this article is organized as follows. We depict the issue of timing errors in Sec.II, as well as the system model. By deriving BMs and CMs, Sec.III illustrates the TEVE algorithm. In Sec.IV, we elaborate the expressions of BER. The upper lower and asymptotic bounds are derived in Sec.IV-A without the TEVE algorithm, while the analytical BER expression of the TEVE algorithm is obtained in Sec.IV-B. In addition, the circumstances of "bad events" are analyzed in Sec.IV-C. The indoor experiment and corresponding results are illustrated in Sec.V. In the end, conclusions are drawn in Sec.VI. It's also mentioned that the variables are illustrated as the lowercase italic forms. Besides, all the vectors in this article are column vectors, which have the lowercase bold forms. The superscript $\bullet \mathbf{T}$ denotes the transpose operation. For ease of reading, the definitions of the main variables are summarized in Appendix B.

## II. Problem Formulation and System Model

In this article, a turbulence-introduced point to point ( P 2 P ) IM/DD FSO link with misalignments is considered. In the OOK scheme, the $k$-th data $x_{k} \in\{0,1\}$ is represented by either the presence ("on") or absence ("off") at the transmitter, and the corresponding energy per bit is described as $\left\{0,2 P_{t} / R_{b}\right\} . P_{t}$ and $R_{b}$ stand for the average transmitting power and the data rate, respectively.

The aggregated channel $h$ is expounded, consisting of path attenuation $h_{l}$, the turbulence fading $h_{a}$, and the pointing error loss $h_{p}$. This article supposes $h_{a}$ has the Gamma-Gamma distribution, while the pointing errors obey the Rayleigh distribution at the receiving plane. Therefore, the probability density function (PDF) of the aggregated channel $h=h_{l} \cdot h_{a} \cdot h_{p}$ is furnished by [24]

$$
\begin{align*}
f_{h}(h)= & \frac{\alpha \beta \rho^{2}}{A_{0} h_{l} \Gamma(\alpha) \Gamma(\beta)} \\
& \cdot \mathbf{G}_{1,3}^{3,0}\left(d \frac{\alpha \beta h}{A_{0} h_{l}} \left\lvert\, \begin{array}{c}
\rho^{2} \\
\rho^{2}-1, \alpha-1, \beta-1
\end{array}\right.\right) \tag{1}
\end{align*}
$$

where $\mathbf{G}(\bullet)$ denotes the Meijer' G function. $\alpha$ and $\beta$ represent the effective numbers of large and small scale turbulent eddies, respectively. $\Gamma(\bullet)$ is the Gamma function. $A_{0}$ denotes the maximum fraction of the collected power in the receiving lens. $\rho=w_{z e q} / 2 \sigma_{s}$ represents the ratio between the equivalent beam radius $w_{z e q}$ and the displacement standard deviation $\sigma_{s}$.

After propagating through the compound channel $h$, the receiver converts the lasers into electrical signals by a photodiode. For an arbitrary positive integer $k$, it's assumed that $\xi_{k}$ stands for the timing error of $k$-th symbol, which means the difference between the edge of sampled clock and the edge of optimal clock, shown in Fig.1. Cheated by the timing errors, the receiver will think the edges of the $k$-th bit are equal to $k-\xi_{k}$ and $k+1-\xi_{k+1}$ in the time axis, which should be $k$ and $k+1$. Then the interference items are introduced by neighboring bits. This article analyzes each timing error $\xi_{k}$ by dividing it into the absolute value $\left|\xi_{k}\right|$ and the sign $\Xi\left(\xi_{k}\right)=+, \xi_{k} \geq 0$ (or $\Xi\left(\xi_{k}\right)=-, \xi_{k}<0$ ). ${ }^{1}$ For ease of analysis, this article normalizes $T_{b}$ to 1 . We also define $\varsigma_{k}$ as the absolute timing errors $\varsigma_{k}=\left|\xi_{k}\right|$. Due to the Gaussian distribution of $\xi_{k}$, the PDF of $\varsigma_{k}$ is illustrated in Eq.(2) [25], [26].

$$
\begin{equation*}
f_{\varsigma}(\varsigma)=\frac{\sqrt{2}}{\sqrt{\pi} \sigma_{\xi}} \exp \left(-\frac{\varsigma^{2}}{2 \sigma_{\xi}^{2}}\right), \varsigma>0 \tag{2}
\end{equation*}
$$

where $\sigma_{\xi}^{2}$ stands for the variance of $\xi_{k}$ (or the second order origin moment of $\varsigma_{k}$ ). Seeing that the unbounded character of $\varsigma_{k}$, the interference of several near bits should be considered, which makes the analyses complicated. In order to discuss the issue of timing errors conveniently, this article defines two complementary events, which are the "clean event" and the "bad event", with the corresponding probabilities of $P_{c l}$ and $P_{b a d}$. The former only takes two adjacent bits into consideration i.e. $x_{k-1}$ and $x_{k+1}$, which is mainly analyzed in this article. It may not be a ridiculous assumption, because it's a high probability event caused by small $\sigma_{\xi}$. For example, the probability of "bad event" is equal to $5.73 \times 10^{-7}$ with $\sigma_{\xi}=0.2$, while it's a much rarer event in $\sigma_{\xi}=0.1$ with the probability equal to $2.21 \times 10^{-16}$. The latter "bad event" treats extra consecutive neighboring $i$ bits as the interference items, where these bits may be detected incorrectly with $\varsigma_{k}$ larger

[^0]than $i$. Thus the average BER $P_{s}$ can be calculated in these two events, given in Eq.(3).
\[

$$
\begin{align*}
P_{s} & =P_{s \mid c l} \cdot P_{c l}+P_{s \mid b a d} \cdot P_{b a d} \\
& \approx P_{s \mid c l}+\sum_{i} i \cdot\left[Q\left(i+1 / \sigma_{\xi}\right)-Q\left(i / \sigma_{\xi}\right)\right] \\
& \approx P_{s \mid c l} \tag{3}
\end{align*}
$$
\]

where $P_{s \mid c l}$ and $P_{s \mid b a d}$ stand for the conditional BER in the "clean events" and "bad events", respectively. $P_{c l}$ and $P_{b a d}$ represent the probabilities of these two events. $Q(\bullet)$ is the $Q$-function denoting $\int_{\bullet}^{+\infty} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{1}{2} t^{2}\right) d t$.

In this sequel, the "clean event" dominates the BER result. Therefore, it is mainly discussed in this article. It needs to be mentioned that the "bad event" is depicted in Sec.IV-C. According to the criterion of "clean event", only the neighboring $(k-1)$-th and $(k+1)$-th bits contribute to the ISI, when decoding any arbitrary $k$-th bit. There are four possible ISI situations decided by different signs of $\xi_{k}$ and $\xi_{k+1}$ accordingly, as depicted in Fig.1. $\Omega_{k}$ is defined as the signs of two continuous timing errors, i.e. $\Omega_{k}=\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)$. Here's an instance. In the case of $\Omega_{k}=++$, there will be $\varsigma_{k}$ portion of $x_{k-1}$ as the interference to $k$-th bit. Besides, only $1-\varsigma_{k+1}$ portion of $x_{k}$ contributes to the $k$-th bit, while the part of $x_{k}$ would become the interference to the $(k+1)$-th bit. In the similar way, the receiving electric signal $\mu_{k}$ with $\Omega_{k}= \pm \pm$ could be written as

$$
\begin{equation*}
\mu_{k}=\eta P_{t} h \mathbf{x}_{k}^{\mathbf{T}} \cdot \boldsymbol{\varsigma}_{k}^{ \pm \pm}+n_{k}, \tag{4}
\end{equation*}
$$

where $\eta$ stands for the photodetector responsivity. $n_{k}$ denotes the equivalent noise with the standard deviation $\sqrt{R_{b}} \sigma_{n} . \mathbf{x}_{k}$ is the vector $\left[x_{k-1}, x_{k}, x_{k+1}\right]$, denoting the $k$-th signal bit and its neighboring interference bits in the "clean events." $\varsigma_{k}^{ \pm \pm}$represents the corresponding weight vector for the signal and interference in four kinds of situations caused by $\Omega_{k}= \pm \pm$. That is,

$$
\begin{align*}
\varsigma_{k}^{++} & =\left[\varsigma_{k}, 1-\varsigma_{k+1}, 0\right] \quad \varsigma_{k}^{--}=\left[0,1-\varsigma_{k}, \varsigma_{k+1}\right] \\
\varsigma_{k}^{+-} & =\left[\varsigma_{k}, 1, \varsigma_{k+1}\right] \quad \varsigma_{k}^{-+}=\left[0,1-\varsigma_{k+1}-\varsigma_{k}, 0\right] \tag{5}
\end{align*}
$$

Owing to neighboring interference, the signal-to-interference-plus-noise ratio (SINR) $\tilde{\gamma}_{ \pm \pm}$is discussed. When $x_{k}$ is equal to $x_{k+1}$ (or $x_{k-1}$ ), the coefficient of $x_{k+1}$ (or $x_{k-1}$ ) can be considered to provide a diversity gain to signal $x_{k}$, which involves the numerator in the SINR expression. However, in the situation of $x_{k+1} \neq x_{k}$ (or $x_{k-1} \neq x_{k}$ ), the coefficient of $x_{k+1}$ (or $x_{k-1}$ ) should be considered as interference, contributing to the denominator in SINR expression. Note that the latter has a dominant influence on degrading the system performance. Thus, this article considers the worst situation with $\tilde{\gamma}_{ \pm \pm}$calculated in Eq.(6).

$$
\begin{align*}
& \tilde{\gamma}_{++}=\frac{\left(1-\varsigma_{k+1}\right) P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\varsigma_{k} P_{t}^{2} h^{2} \eta^{2}} \tilde{\gamma}_{--}=\frac{\left(1-\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\varsigma_{k+1} P_{t}^{2} h^{2} \eta^{2}} \\
& \tilde{\gamma}_{+-}=\frac{P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\left(\varsigma_{k+1}+\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}} \\
& \tilde{\gamma}_{-+}=\frac{\left(1-\varsigma_{k+1}-\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}} \tag{6}
\end{align*}
$$



Fig. 1. The diagrammatic sketch of timing errors and four cases of continuous timing errors $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=++,-+,+-,--$.

From Eq.(6), it's obtained that the $\operatorname{SINR} \tilde{\gamma}$ has an asymptotic value, also defined as $\tilde{\gamma}_{ \pm \pm}^{\infty}=\lim _{P_{t} \rightarrow \infty} \tilde{\gamma}_{ \pm \pm}$, which is deduced in Eq.(7).

$$
\begin{array}{ll}
\tilde{\gamma}_{++}^{\infty}=\frac{1-\varsigma_{k+1}}{\varsigma_{k}} & \tilde{\gamma}_{--}^{\infty}=\frac{1-\varsigma_{k}}{\varsigma_{k+1}} \\
\tilde{\gamma}_{+-}^{\infty}=\frac{1}{\varsigma_{k+1}+\varsigma_{k}} & \tilde{\gamma}_{-+}^{\infty}=\left(1-\varsigma_{k}-\varsigma_{k+1}\right) \gamma^{\infty}, \tag{7}
\end{array}
$$

where $\gamma^{\infty}=P_{t}^{2} h^{2} \eta^{2} / R_{b} \sigma_{n}^{2},\left(P_{t} \rightarrow \infty\right)$ is defined as the infinity SNR (signal-to-noise ratio) without timing errors. It can be maintained that timing errors hamper the system performance. It's urgent to mitigate the impairment. Thus the TEVE algorithm is illustrated in the next section.

## III. TEVE Algorithm

In this section, the TEVE algorithm will be introduced. Although the TEVE algorithm is designed under the circumstances of "clean events", it can also provide some help to the case of "bad events", as one can obtain from Sec.IV-C. As mentioned above, for decoding the $k$-th signal, the neighboring $(k \pm 1)$-th bits would be considered as the interference items with the weights of $\varsigma_{k}^{ \pm \pm}$in the case of "clean events". Thus the MLSD scheme is utilized to removing the interference and decoding the signal bit $\hat{x}_{k}$ for any integer $k \geq 2$, which is

$$
\begin{align*}
\hat{x}_{k} & =\underset{x_{k}}{\arg \max } \sum_{q=1}^{k+1} \ln \left[p^{ \pm \pm}\left(\mu_{q} \mid \mathbf{x}_{q}, \varsigma_{q}^{ \pm \pm}, h\right)\right] \\
& =\underset{x_{k}}{\arg \min } \sum_{q=1}^{k+1} \lambda_{q}^{ \pm \pm}\left(\mu_{q} \mid \mathbf{x}_{q}, \varsigma_{q}^{ \pm \pm}, h\right) \tag{8}
\end{align*}
$$

where $p^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \boldsymbol{\varsigma}_{k}^{ \pm \pm}, h\right)$ means the conditional PDF with known $\boldsymbol{\varsigma}_{k}^{ \pm \pm}$and $h$, and it's equal to $\frac{1}{\sqrt{2 \pi} \sigma} \cdot \exp \left(-\frac{1}{2 \sigma^{2}} \cdot\left(\mu_{k}-h \boldsymbol{\varsigma}_{k}^{ \pm \pm} \cdot \mathbf{x}_{k}\right)^{2}\right) . \quad$ Besides, $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \boldsymbol{\varsigma}_{k}^{ \pm \pm}, h\right)=\left(\mu_{k}-h \boldsymbol{\varsigma}_{k}^{ \pm \pm} \cdot \mathbf{x}_{k}\right)^{2}$ denotes the logarithm operation results of $p^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \boldsymbol{\varsigma}_{k}^{ \pm \pm}, h\right)$ and discards the constant terms.

Owing to the fact that the channel gain $h$ and timing errors ( $\varsigma_{k}$ and $\varsigma_{k+1}$ ) are random variables, this article supposes that the receiver doesn't obtain the instantaneous values of $h$ and $\varsigma$,
but has the knowledge of their PDFs. The PDF of the channel can be estimated by parameters estimation [44], while the PDF of timing errors can be obtained by sampling the clock at a higher frequency or by the oscilloscope. Therefore, the BM $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ should be calculated by the mathematical expectation of $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \boldsymbol{\varsigma}_{k}^{ \pm \pm}, h\right)$, which is

$$
\begin{align*}
\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)= & \mathbb{E}_{h, \boldsymbol{\varsigma}_{k}^{ \pm \pm}}\left[\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \boldsymbol{\varsigma}_{k}^{ \pm \pm}, h\right)\right] \\
= & \left.\iiint \int_{f_{\varsigma}}\left(\mu_{k}-h \varsigma_{k}^{ \pm \pm}\right) d \varsigma_{k} \cdot f_{\varsigma}\left(\varsigma_{k+1}\right) d\right)^{2} f_{h}(h) d h \\
& \cdot \varsigma_{k+1} \tag{9}
\end{align*}
$$

For every BM in the $k$-th time slot, the common item $\mathbb{E}_{h, \varepsilon_{k}^{ \pm \pm}}\left[u_{k}^{2}\right]$ from Eq.(9) can be ignored, because it makes no contribution to comparing BMs and CMs. During the simplification process of Eq.(9), three mathematical expectations are needed for the variables $\varsigma_{k}, \varsigma_{k+1}$ and $h$. Eq.(10) and Eq.(11) define two intermediate results $\lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{ \pm \pm}$and $\lambda_{k, \mathbb{E}\left[\varsigma_{k}, \varsigma_{k+1}\right]}^{ \pm \pm}$, respectively, which can help to derive the closed-forms of BMs in Eq.(9).

$$
\begin{align*}
\lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{ \pm \pm} & =\mathbb{E}_{\varsigma_{k}}\left[\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \varepsilon_{k}^{ \pm \pm}, h\right)\right]  \tag{10}\\
\lambda_{k, \mathbb{E}\left[\varsigma_{k}, \varsigma_{k+1}\right]}^{ \pm \pm} & =\mathbb{E}_{\varsigma_{k}, \varsigma_{k+1}}\left[\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}, \varepsilon_{k}^{ \pm \pm}, h\right)\right] \tag{11}
\end{align*}
$$

After the integral operation of $\varsigma_{k}$ in Eq.(9), the results of $\lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{ \pm \pm}$are derived in Eq.(12) with $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)= \pm \pm$.

$$
\begin{aligned}
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{++} \\
&= \frac{\eta h P_{t}}{2 \sqrt{2 \pi}} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right)\left[-2 \sqrt{\frac{2}{\pi}} x_{k-1}\right. \\
& \cdot\left(\eta h P_{t} \varsigma_{k+1} x_{k}-\eta h P_{t} x_{k}+\mu_{k}\right)+\frac{1}{\sigma_{\xi}} x_{k}\left(\varsigma_{k+1}-1\right) \\
&\left.\cdot\left(\eta h P_{t} \varsigma_{k+1} x_{k}-\eta h P_{t} x_{k}+2 \mu_{k}\right)+\eta h P_{t} \sigma_{\xi} x_{k-1}^{2}\right] \\
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{+-} \\
&= \frac{\eta h P_{t}}{2 \sqrt{2 \pi}} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right)\left[2 \sqrt{\frac{2}{\pi}} x_{k-1}\right. \\
& \cdot\left(\eta h P_{t} \varsigma_{k+1} x_{k+1}+\eta h P_{t} x_{k}-\mu_{k}\right) \\
&+\frac{1}{\sigma_{\xi}}\left(\varsigma_{k+1} x_{k+1}+x_{k}\right)
\end{aligned}
$$

$$
\begin{align*}
&\left.\cdot\left(\eta h P_{t} s_{k+1} x_{k+1}+\eta h P_{t} x_{k}-2 \mu_{k}\right)+\sigma_{\xi} \eta h P_{t} x_{k-1}^{2}\right] \\
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}^{-+} \\
&= \frac{\eta h P_{t}}{2 \sqrt{2 \pi}} x_{k} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right) \\
& \cdot\left[2 \sqrt{\frac{2}{\pi}}\left(\eta h P_{t} \varsigma_{k+1} x_{k}-\eta h P_{t} x_{k}+\mu_{k}\right)+\frac{\left(\varsigma_{k+1}-1\right)}{\sigma_{\xi}}\right. \\
&\left.\cdot\left(\eta h P_{t} \varsigma_{k+1} x_{k}-\eta h P_{t} x_{k}+2 \mu_{k}\right)+\sigma_{\xi} \eta h P_{t} x_{k}\right] \\
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}\right]}= \frac{\eta h P_{t}}{2 \sqrt{2 \pi}} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right) \\
& \cdot\left[-2 \sqrt{\frac{2}{\pi}} x_{k}\left(\eta h P_{t} \varsigma_{k+1} x_{k+1}+\eta h P_{t} x_{k}-\mu_{k}\right)\right. \\
& {\left[\frac{1}{\sigma_{\xi}}\left(s_{k+1} x_{k+1}+x_{k}\right)\right.} \\
&\left.\cdot\left(\eta h P_{t} \varsigma_{k+1} x_{k+1}+\eta h P_{t} x_{k}-2 \mu_{k}\right)+\sigma_{\xi} \eta h P_{t} x_{k}^{2}\right]
\end{align*}
$$

Similarly, the mathematical average results of Eq.(12) on $\varsigma_{k+1}$ are equal to $\lambda_{k, \mathbb{E}\left[\varsigma_{k} \varsigma_{k+1}\right]}^{ \pm \pm}$, which is deduced in Eq.(13). It's noted that the common efficient $h$ in Eq.(13) cannot be dropped, because $\mathbb{E}\left[h^{2}\right]$ doesn't have a square relationship with $\mathbb{E}[h]$.

$$
\begin{align*}
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}, \varsigma_{k+1}\right]}^{++} \\
& =\frac{\eta h P_{t}}{4 \pi}\left[\eta h P _ { t } \sigma _ { \xi } ^ { 2 } \left(\pi x_{k-1}^{2}-4 x_{k-1} x_{k}\right.\right. \\
& \left.+\pi x_{k}^{2}\right)+\pi x_{k}\left(-2 \mu_{k}+\eta h P_{t} x_{k}\right) \\
& \left.+2 \sqrt{2 \pi} \sigma_{\xi}\left(x_{k-1}-x_{k}\right)\left(-\mu_{k}+\eta h P_{t} x_{k}\right)\right] \\
& \lambda_{k, \mathbb{E}\left[\zeta_{k}, \zeta_{k+1}\right]}^{+-} \\
& =\frac{\eta h P_{t}}{4 \pi}\left[\pi \eta h P_{t} \sigma_{\xi}^{2}\left(x_{k-1}^{2}+x_{k+1}^{2}\right)\right. \\
& +\pi x_{k}\left(-2 \mu_{k}+\eta h P_{t} x_{k}\right)+4 \sigma_{\xi}^{2} \eta h P_{t} x_{k-1} x_{k+1} \\
& \left.-2 \sqrt{2 \pi} \sigma_{\xi}\left(\mu_{k}-\eta h P_{t} x_{k}\right)\left(x_{k-1}+x_{k+1}\right)\right] \\
& \lambda_{k, \mathbb{E}\left[\zeta_{k}, \zeta_{k+1}\right]}^{+} \\
& =\frac{-\eta h P_{t}}{4 \pi} x_{k}\left[\eta h P_{t}\right. \\
& \cdot\left(-(4+2 \pi) \sigma_{\xi}^{2}-\pi+4 \sqrt{2 \pi} \sigma_{\xi}\right) x_{k} \\
& \left.+\left(2 \pi-4 \sqrt{2 \pi} \sigma_{\xi}\right) \mu_{k}\right] \\
& \lambda_{k, \mathbb{E}\left[\varsigma_{k}, \varsigma_{k+1}\right]}^{--} \\
& =\eta h P_{t}\left[\frac{1}{4} x_{k}\left(\eta h P_{t} x_{k}-2 \mu_{k}\right)\right. \\
& +\frac{\sigma_{\xi}}{\sqrt{2 \pi}}\left(x_{k}-x_{k+1}\right)\left(\mu_{k}-\eta h P_{t} x_{k}\right) \\
& +\frac{1}{4} \sigma_{\xi}^{2} \eta h P_{t} x_{k}^{2}-\frac{1}{\pi} \sigma_{\xi}^{2} \eta h P_{t} x_{k+1} x_{k} \\
& \left.+\frac{1}{4} \sigma_{\xi}^{2} \eta h P_{t} x_{k+1}^{2}\right] \tag{13}
\end{align*}
$$

TABLE I
The Pseudo Code Diagram of the TEVE Algorighm

```
Initialization: \(\Lambda_{1}=0,\left\{\mho_{1}\right\}=\{0,0,+\}, \hat{x}_{k-1}=0\).
for \(k=1\) : \(L\)
    Calculate all the next possible state,
    \(\left\{\mho_{k+1}\right\}=\left\{\left\{\mho_{k}(2)\right\}, x_{k+1}=0,1, \Xi\left(\xi_{k+1}\right)\right\}\).
    Calculate all the possible BMs by Eq.(14),
    \(\lambda_{k}^{\mho_{k}(3) \mho_{k+1}(3)}\left\{\mu_{k} \mid \mathbf{x}_{k}=\left[\left\{\mho_{k-1}(1)\right\},\left\{\mho_{k}(1)\right\},\left\{\mho_{k}(2)\right\}\right\}\right)\).
    Calculate all the possible paths,
    \(\Lambda_{k+1}^{t e m p}=\Lambda_{k}+\lambda_{k}^{\mho_{k}(3) \mho_{k+1}(3)}\left(\mu_{k} \mid \mathbf{x}_{k}=\left(\hat{x}_{k-1}, \hat{x}_{k}, 0\right.\right.\) or 1\(\left.)\right)\).
    if \(k \geq \delta+1\)
        make the decision on \(\hat{x}_{k+1-\delta}\) by Eq.(15),
        \(\hat{x}_{k+1-\delta}=\min _{x_{k+1-\delta}}\left(\Lambda_{k+1}^{\text {temp }}\right)\).
        Retain the paths including \(\left\{\mho_{k+1-\delta}\right\}=\left\{\hat{x}_{k-\delta}, \hat{x}_{k+1-\delta}\right.\),
        \(\left.\Xi\left(\xi_{k+1-\delta}\right)\right\}\) and discard the other paths.
    end if
end for
Output: The symbols after decision \(\left\{\hat{x}_{k}\right\}\).
```

By substituting $h$ and $h^{2}$ by the expressions of their mathematical expectation in Eq.(13) and removing common coefficients, the closed form of BMs in Eq.(9) are deduced, in Eq.(14), as shown at the bottom of the next page.

After deriving the expressions of BMs, we also define $\Lambda_{k}$ as the CM in the $k$-th time unit for any positive integer $k$, which is equal to the sum of the previous CM $\Lambda_{k-1}$ and the current $\mathrm{BM} \lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$. The initial value $\Lambda_{1}$ is set to be 0 . According to the add-compare-select iteration, $\Lambda_{k}^{\text {temp }}$ is assumed to be the possible CM before selection. On the basis of sliding window decoder, we choose the decoding delay $\delta$ which is much larger than 2 . When we have received the $(k+\delta)$-th bit, we can backtrack for $\delta$ steps. The $k$-th bit could be decoded by choosing $\hat{x}_{k}$ which leads to the smallest CM $\Lambda_{k+\delta}$, furnished by Eq.(15). For brevity, the flowchart of the TEVE method is shown in Table I by a pseudo code diagram.

$$
\begin{equation*}
\hat{x}_{k}=\min _{x_{k}}\left(\Lambda_{k+\delta-1}+\lambda_{k+\delta}^{ \pm \pm}\left(u_{k+\delta} \mid \mathbf{x}_{k+\delta}\right)\right) \tag{15}
\end{equation*}
$$

Note that the differences between the TEVE algorithm and the traditional Viterbi method are illustrated as follows. First, the number of states becomes double, owing to the signs $\Xi\left(\xi_{k}\right)=+$ or - . Besides, the BM has the form of expectation combined of $h, \varsigma_{k}$ and $\varsigma_{k+1}$. For each time interval, there are 8 states. We may define $\mho_{k}=\left\{x_{k-1}, x_{k}, \Xi\left(\xi_{k}\right)\right\}$ as the correct state in the $k$-th time interval, in view of the transmitted signal $x_{k-1}$ and $x_{k}$. We may suppose the $\mho_{k}(i)$ to refer the $i$-th $(i=1,2,3)$ element in $\mho_{k}(i)$, i.e. $\mho_{k}(1)=$ $x_{k-1}, \mho_{k}(2)=x_{k}, \mho_{k}(3)=\Xi\left(\xi_{k}\right)$. Similarly, we define $\hat{\mho}_{k}=\left\{\hat{x}_{k-1}, \hat{x}_{k}, \Xi\left(\xi_{k}\right)\right\}$ as the state of the $k$-th in the TEVE algorithm after decision, while $\left\{\mho_{k}\right\}=\left\{x_{k-1}, x_{k}, \Xi\left(\xi_{k}\right)\right\}$ denotes all the possible states for the $k$-th time interval. The trellis diagram of TEVE algorithm is depicted in Fig.2.

As seen in Fig.2, the left side picture illustrates the possible paths between two adjacent bits. In Fig.2, blue points stand for $\Xi\left(\xi_{k}\right)=-$, while the red points denote the other case $\Xi\left(\xi_{k}\right)=+$. The right side picture indicates the example paths. For the $k$-th time interval, every arrow represents a possible path which maps a relevant branch metric $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$. It's seen from Fig. 2 that there will always be four paths between the arbitrary two states $\mho_{k}$ and $\mho_{k+3}$, which results from the situations of different bits ( 0 or 1 ) and different signs ( + or - )


Fig. 2. Possible paths between neighboring states (Left) and an example trellis diagram of the TEVE algorithm (Right).
of corresponding timing errors. From Fig.2, it's apparent to see there are also two paths for arbitrary two separated states $\mho_{k-1}$ and $\mho_{k+1}$. These two paths with the same starting and ending points can form a quadrilateral, and we name it "girth4". Benefitting from the "girth-4", the TEVE algorithm can still make correct detections on the bit $\hat{x_{k}}$ even though path with wrong $\Xi\left(\xi_{k}\right)$ is remained. Because the two paths in the "girth-4" of $\mho_{k-1}$ and $\mho_{k+1}$ have the common $x_{k}$ and $x_{k+1}$. For example, it can be indicated that both $x_{4}$ and $x_{5}$ are equal to 1 in the marked "girth-4" from Fig.2.

## IV. Performance Analysis

In this section, we first evaluate the BER expressions with timing errors in Sec.IV-A, while the BER of TEVE algorithm will be depicted in Sec.IV-B. In the first subsection, we deduce the upper and lower bounds ( $P_{s}^{u p}$ and $P_{s}^{l o w}$ ) of BER, as well as the asymptotic bound $P_{s}^{\infty}$. In the second subsection, we derive the approximate mathematic form of $P_{s}^{T E V E}$. Both the two subsections discuss the circumstance of "clean events". In the Sec.IV-C, "bad events" are analyzed.

## A. Analysis on Timing Errors' Impairment

As illustrated in Sec.II, the average BER $P_{s}$ is approximate to the mean BER $P_{s, c l}$ of the "clean events." Obeying

Gaussian distribution, any timing error $\xi_{k}$ is supposed to have equal probability of $\Xi\left(\xi_{k}\right)=+$ and $\Xi\left(\xi_{k}\right)=-$. Therefore, the mean BER $P_{s}$ could be derived as

$$
\begin{align*}
P_{s} \approx & P_{s, c l}=\frac{1}{4} \sum_{i \in \Omega_{k}} \iiint Q\left(\tilde{\gamma}_{i}\right) \cdot \frac{\alpha \beta \rho^{2}}{A_{0} \cdot h_{l} \Gamma(\alpha) \Gamma(\beta)} \\
& \cdot \mathbf{G}_{1,3}^{3,0}\left(\left.\frac{\alpha \beta h}{A_{0} \cdot h_{l}} \right\rvert\, \rho^{2} ; \rho^{2}-1, \alpha-1, \beta-1\right) d h \\
& \cdot \frac{\sqrt{2}}{\sqrt{\pi} \sigma_{\xi}} \exp \left(-\frac{\varsigma_{k}^{2}}{2 \sigma_{\xi}^{2}}\right) d \varsigma_{k} \frac{\sqrt{2}}{\sqrt{\pi} \sigma_{\xi}} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right) d \varsigma_{k+1} \tag{16}
\end{align*}
$$

It can be seen from Eq.(16) that the BER $P_{s, c l}$ is made up of four situations. We may mark them as $P_{s}^{++}, P_{s}^{+-}, P_{s}^{-+}, P_{s}^{+-}$, according to the corresponding $\Omega_{k}$. Fig. 3 depicts the BER simulation of the four cases, utilizing the simulation factor in Table III of Sec.IV. As obtained in Fig.3, the worst case is $\Omega_{k}=+-$. In this case, the $k$-th bit suffers from the interferences of $(k-1)$-th bit and $(k+1)$-th bit. In the case of $\Omega_{k}=--$ (or ++ ), there will be one neighboring symbol contributing the interference item, i.e. $(k+1)$-th bit (or $k$-th bit). At the same time, the portion $1-\varsigma_{k}$ (or $1-\varsigma_{k+1}$ ) of the signal bit remains. Thus the two cases have almost the same performance and have a better performance than

$$
\begin{align*}
\lambda_{k}^{++}\left(\mu_{k} \mid \mathbf{x}_{k}\right)= & \frac{\eta P_{t}(\alpha+1) A^{2}(\beta+1) \rho^{2}}{\pi \alpha \beta\left(\rho^{2}+2\right)} \cdot\left[\pi \sigma_{\xi}^{2} x_{k-1}^{2}-4 \sigma_{\xi}^{2} x_{k-1} x_{k}+2 \sqrt{2 \pi} \sigma_{\xi}\left(x_{k-1}-x_{k}\right) x_{k}+\pi\left(1+\sigma_{\xi}^{2}\right) x_{k}^{2}\right] \\
& -\frac{A_{0} \rho^{2}}{\pi\left(1+\rho^{2}\right)}\left[2 \sqrt{2 \pi} \sigma_{\xi} \mu_{k}\left(x_{k-1}-x_{k}\right)-2 \pi \mu_{k} x_{k}\right] \\
\lambda_{k}^{+-}\left(\mu_{k} \mid \mathbf{x}_{k}\right)= & \frac{\eta h P_{t}(\alpha+1) A^{2}(\beta+1) \rho^{2}}{\alpha \beta\left(\rho^{2}+2\right)} \cdot\left[\sigma_{\xi}^{2} x_{k-1}^{2}+x_{k}^{2}+\frac{4 \sigma_{\xi}^{2}}{\pi} x_{k-1} x_{k+1}+\sigma_{\xi}^{2} x_{k+1}^{2}+\frac{2 \sqrt{2 \pi} \sigma_{\xi}}{\pi} x_{k}\left(x_{k-1}+x_{k+1}\right)\right] \\
& -\frac{A_{0} \rho^{2}}{1+\rho^{2}}\left[\frac{2 \sqrt{2 \pi} \sigma_{\xi}}{\pi} \mu_{k}\left(x_{k-1}+x_{k+1}\right)-2 \mu_{k} x_{k}\right] \\
\lambda_{k}^{-+}\left(\mu_{k} \mid \mathbf{x}_{k}\right)= & \frac{\eta h P_{t}(\alpha+1) A^{2}(\beta+1) \rho^{2}}{\alpha \beta\left(\rho^{2}+2\right) \pi}\left(\pi-4 \sqrt{2 \pi} \sigma_{\xi}+4 \sigma_{\xi}^{2}+2 \pi \sigma_{\xi}^{2}\right) x_{k}^{2}+\frac{2 A_{0} \rho^{2}}{\pi\left(1+\rho^{2}\right)}\left(-\pi+2 \sqrt{2 \pi} \sigma_{\xi}\right) \mu_{k} x_{k} \\
\lambda_{k}^{--}\left(\mu_{k} \mid \mathbf{x}_{k}\right)= & \frac{\eta h P_{t}(\alpha+1) A^{2}(\beta+1) \rho^{2}}{\alpha \beta\left(\rho^{2}+2\right)} \cdot\left[x_{k}^{2}+\sigma_{\xi}^{2} x_{k}^{2}-\frac{2 \sqrt{2 \pi}}{\pi} \sigma_{\xi} x_{k}\left(x_{k}-x_{k+1}\right)-\frac{4}{\pi} \sigma_{\xi}^{2} x_{k+1} x_{k}+\sigma_{\xi}^{2} x_{k+1}^{2}\right] \\
& +\frac{A_{0} \rho^{2}}{1+\rho^{2}}\left[\frac{2 \sqrt{2 \pi}}{\pi} \sigma_{\xi} \mu_{k}\left(x_{k}-x_{k+1}\right)-2 \mu_{k} x_{k}\right] \tag{14}
\end{align*}
$$



Fig. 3. BER results of four cases $\left(P_{s}^{++}, P_{s}^{-+}, P_{s}^{+-}, P_{s}^{--}\right)$versus transmitting power.
the case of $\Omega_{k}=+-$. In the last case of $\Omega_{k}=-+$, the timing errors only lead to the loss portion of $\varsigma_{k}+\varsigma_{k+1}$ in the $k$-th signal and no interference from other bits. That's why $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=-+$ could be seen as the most friendly case, which causes the ideal curve moving towards the positive direction with the coefficient of $1 / 1-\varsigma_{k}-\varsigma_{k+1}$. In addition, it's apparent that the BER results increase with larger $\sigma_{\xi}$.

However, Eq.(16) is beyond the authors' ability to simplify. As an alternative, we try to give the upper and lower bounds of Eq.(16). We discuss the condition that the transmitting power is large enough such that the SINR $\tilde{\gamma}_{ \pm \pm}$could be an improper fraction, i.e. $\tilde{\gamma}>1$. In this case, $\tilde{\gamma}_{+-}$becomes the smallest value of the set of $\tilde{\gamma}_{ \pm \pm}$, i.e. $\tilde{\gamma}_{+-}=\frac{P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\left(\varsigma_{k-1}+\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}}<$ $\tilde{\gamma}_{++}, \tilde{\gamma}_{-+}, \tilde{\gamma}_{--}$. Because $\tilde{\gamma}_{+-}$is able to be considered as adding an identical value in both the denominator and the numerator of any other SINR. For example, $\tilde{\gamma}_{++}=$ $\frac{\left(1-\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\varsigma_{k-1} P_{t}^{2} h^{2} \eta^{2}} \geq \frac{\left(1-\varsigma_{k}\right) P_{t}^{2} h^{2} \eta^{2}+\varsigma_{k} P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\varsigma_{k-1} P_{t}^{2} h^{2} \eta^{2}+\varsigma_{k} P_{t}^{2} h^{2} \eta^{2}}=\tilde{\gamma}_{+-}$. Then the lower bound $P_{s}^{\text {low }}$ could be derived by reserving the $\Omega_{k}=+-$ case and discarding other cases. With the help of Jensen's inequality, $P_{s}^{l o w}$ could be further simplified as

$$
\begin{align*}
& P_{s}^{l o w} \\
&= \iiint Q\left(\tilde{\gamma}_{+-}\right) f_{h}(h) d h \\
& \cdot f_{\varsigma}\left(\varsigma_{k}\right) d \varsigma_{k} \cdot f_{\varsigma}\left(\varsigma_{k+1}\right) d \varsigma_{k+1} \\
& \iiint Q\left(\frac{P_{t}^{2} \mathbb{E}\left[h^{2}\right] \eta^{2}}{R_{b} \sigma_{n}^{2}+\left(s_{k+1}+\varsigma_{k}\right) P_{t}^{2} \mathbb{E}\left[h^{2}\right] \eta^{2}}\right) \\
& \cdot \frac{\sqrt{2}}{\sqrt{\pi} \sigma_{\xi}} \exp \left(-\frac{\varsigma_{k}^{2}}{2 \sigma_{\xi}^{2}}\right) \cdot d \varsigma_{k} \frac{\sqrt{2}}{\sqrt{\pi} \sigma_{\xi}} \exp \left(-\frac{\varsigma_{k+1}^{2}}{2 \sigma_{\xi}^{2}}\right) \cdot d \varsigma_{k-1} \tag{17}
\end{align*}
$$

By adopting Gaussian-Hermite polynomials [31], the closed form expression of Eq.(17) is furnished in Eq.(18), as shown at the bottom of the next page, where $\omega_{i}$ and $z_{i}$ denote the roots and the weights of $i$-th-order Hermite polynomials, respectively.

Noting the fact that $\tilde{\gamma}_{+-}<\tilde{\gamma}_{++}, \tilde{\gamma}_{-+}, \tilde{\gamma}_{-+}$in large $P_{t}$ situation, the upper bound $P_{s}^{u p}$ could be obtained by replacing $\tilde{\gamma}_{++}, \tilde{\gamma}_{-+}, \tilde{\gamma}_{-+}$with $\tilde{\gamma}_{+-}$. The upper bound $P_{s}^{u p}$ in $\mathrm{Eq}(19)$, as shown at the bottom of the next page.

In the ideal optical wireless communication link without timing errors, it is apparent that the BER $P_{s}$ enhances with


Fig. 4. The upper bound $P_{s}^{u p}$ (dash-dotted lines), lower bound $P_{s}^{l o w}$ (dashdotted lines) and asymptotic bound $P_{s}^{\infty}$ for average BER $P_{s}$ (solid lines and markers).
the higher SNR $\gamma$. There seems to be no limit that restricts the performance. However, the timing error $\xi$ provides an asymptotic value even with enormous SNR $\gamma$, which is defined as the asymptotic bound $P_{s}^{\infty}=\lim _{P_{t} \rightarrow \infty} P_{s} \approx \lim _{P_{t} \rightarrow \infty} P_{s, c l}$. The asymptotic bound $P_{s}^{\infty}$ could be derived from Eq.(7), which is given in Eq. (20), as shown at the bottom of the next page.

The asymptotic bound $P_{s}^{\infty}$ indicates that there will be an error floor introduced by $\sigma_{\xi}$. Fig. 4 analyzes the BER results and corresponding bounds. As revealed in Fig.4. the lower bound $P_{s}^{l o w}$ is confirmed to be tight to the actual $P_{s}$ with the increasing transmitting power $P_{t}$. The actual $P_{s}$ approaches the asymptotic boundary $P_{s}^{\infty}$, as well as the lower bound $P_{s}^{l o w}$, which confirms the error floor conclusion by Eq. $(7,20)$. Paralleling to $P_{s}^{l o w}$, the upper bound $P_{s}^{u p}$ keeps a certain distance from the actual BER $P_{s}$. Noting the curve distances of different $\sigma_{\xi}$, it could be conclude that small $\sigma_{\xi}$ may be tolerable, while large $\sigma_{\xi}$ impairs the system performance badly.

## B. Analysis on TEVE

In order to measure the performance by TEVE algorithm, the theoretical BER performance is deduced in this subsection. We suppose that the TEVE algorithm has made the wrong decision on the $(k+1)$-th bit, but would return to the correct path at $(k+l)$-th interval. The error event is defined as $\Theta_{l}$, which is $\Theta_{l}=\left\{\hat{\mho}_{k}=\mho_{k}\right\} \&\left\{\hat{\mho}_{k+l}=\mho_{k+l}\right\} \&\left\{\hat{\mho}_{k+i}=\right.$ $\left.\mho_{k+i}, i=1,2, \ldots, l-1\right\}$. It's explicit that $l$ is larger than or equal to 3 , according to Fig.3. The error event $\Theta_{l}$ results from the condition that the wrong path's CM is larger than that of right path, whose conditional probability $P\left(\Theta_{l} \mid h\right)$ is furnished in Eq.(21).

$$
\begin{align*}
P\left(\Theta_{l} \mid h\right) \approx & P_{c l}\left(\Theta_{l} \mid h\right) \\
= & P\left[\sum_{i=k}^{k+l-1}\left(\mu_{k}-\eta P_{t} \varsigma_{k} \cdot \hat{\mathbf{x}}_{k} h\right)^{2}\right. \\
& \left.<\sum_{i=k}^{k+l-1}\left(\mu_{k}-\eta P_{t} \varsigma_{k} \cdot \mathbf{x}_{k} h\right)^{2}\right] \tag{21}
\end{align*}
$$

where $P_{c l}\left(\Theta_{l} \mid h\right)$ denotes the conditional probability of error event $\Theta_{l}$ in the case of "clean events". Conceiving of $\mu_{k}=$ $\eta P_{t} h \mathbf{x}_{k}^{\mathbf{T}} \cdot \boldsymbol{\varsigma}_{k}^{ \pm \pm}+n_{k}$ given in Eq.(4), Eq.(21) becomes

$$
\begin{equation*}
P\left(\Theta_{l} \mid h\right) \approx P\left[h<-\frac{2 / \eta P_{t} \sum_{i=k}^{k+l-1} n_{i} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}}{\left[\sum_{i=k}^{k+l-1} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}\right]^{2}}\right] \tag{22}
\end{equation*}
$$

where the vector $\varepsilon_{i}=\left[\varepsilon_{i-1}, \varepsilon_{i}, \varepsilon_{i+1}\right]$ denotes the error difference between $\mathbf{x}_{i}$ and $\hat{\mathbf{x}}_{i}$. The expression of $\boldsymbol{\varsigma}_{i}^{\mathbf{T}}$ is furnished by Eq.(23).

$$
\boldsymbol{\varsigma}_{i}^{\mathbf{T}}=\left(\begin{array}{ccc}
\varsigma_{i}, 1-\varsigma_{i+1} & , & 0  \tag{23}\\
\varsigma_{i}, & 1 & , \\
\varsigma_{i+1} \\
0,1-\varsigma_{i+1} & -\varsigma_{i}, 0 \\
0, & 1-\varsigma_{i} & , \varsigma_{i+1}
\end{array}\right), \begin{aligned}
& , \Omega_{k}=++ \\
& , \Omega_{k}=+- \\
& , \Omega_{k}=-+ \\
& , \Omega_{k}=--
\end{aligned}
$$

Let us focus on the fraction in the brackets in Eq.(22). Owing to that $n_{i}$ being the coefficient of $\varepsilon_{i}$ in the molecule and $\varepsilon_{i}$ has the square form in the denominator, we just need to ponder the absolute value of $\varepsilon_{i}$. We define $\varepsilon_{l}^{e r}=$ $\left[\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \varepsilon_{k+l}\right]$ to be the error vector denoting the error event $\Theta_{l} . \wp_{l}$ is defined as $\wp_{l}=\sum_{i=1}^{l} \varepsilon_{k+i}$, representing the number of " 1 " in $\varepsilon_{l}^{e r}$. Note that $\wp_{l}$ is smaller than $l-1$. As stated above, it's assumed that wrong decision has been made to the $(k+1)$-th bit, i.e. $\varepsilon_{k+1}=1$, due to deviation from the correct path at the $k$-th time interval. It's obtained that both $\varepsilon_{k+2}$ and $\varepsilon_{k+3}$ are equal to 0 , if the path moves back to the $(k+1)$-th time interval. In the similar way, we can reveal the general property of $\varepsilon_{l}^{e r}$, shown as REMARK 1.

Remark 1: In the vector $\varepsilon_{l}^{e r}, \varepsilon_{k+1}$ is always equal to 1 , and $\left[\varepsilon_{k+l-2}, \varepsilon_{k+l-1}, \varepsilon_{k+l}\right]$ satisfies the condition $=[1,0,0]$. For any vectors truncated from $\varepsilon_{l}^{e r}$ containing $\varepsilon_{k+1}$, they cannot have the same sequences as $\left.\varepsilon_{i}^{e r}\right|_{i<l}$.

Lemma 1: For any positive integer $l, \sum_{i=k}^{k+l-1} \varsigma_{i}^{\mathbf{T}} \cdot \varepsilon_{i}$ is always equal to a constant number $\wp_{l}$, and has no relationship with $\varepsilon_{l}^{e r}$.

Proof: For keeping a logic consistency, LEMMA 1 will be proved by mathematical induction, shown in the Appendix A.

In the light of the sophisticated fraction, Eq.(22) is too hard to be simplified. LEMMA 1 helps us to simplify Eq.(22) by replacing the denominator with $\wp_{l}$. The conditional


Fig. 5. Conditional probability $P\left(\Theta_{l} \mid h\right)(l=3,4,5)$ versus normalized channel gain $h$ with $\sigma_{\xi}=0.1$ (solid lines and round markers) and $\sigma_{\xi}=0.2$ (dashed lines and hexagon markers).
expectation $P\left(\Theta_{l} \mid h\right)$ is derived by Eq.(24).

$$
\begin{align*}
& P\left(\Theta_{l} \mid h\right) \\
& \approx \mathbb{E}_{n_{i}, \varsigma_{i}}\left[P\left(h<-2 / \wp \wp_{l} \eta P_{t} \sum_{i=k}^{k+l} n_{i} \varsigma^{\mathbf{T}}{ }_{i} \cdot \varepsilon_{i}\right)\right] \\
&= \frac{A_{0}}{2 \alpha \beta} \mathbb{E}_{n_{i}, \varsigma_{i}}\left[\mathbf { G } _ { 2 , 4 } ^ { 3 , 1 } \left(-\frac{2 \alpha \beta}{A_{0} \eta P_{t} \wp_{l}}\right.\right. \\
&\left.\left.\cdot \sum_{i=k}^{k+2} n_{i} \varsigma^{\mathbf{T}}{ }_{i} \cdot \varepsilon_{i} \left\lvert\, \begin{array}{l}
1,1+\rho^{2} \\
\alpha, \beta, \rho^{2}, 0
\end{array}\right.\right)\right] \\
&= \frac{A_{0}}{\alpha \beta} \cdot \frac{2^{n_{l}-1}}{\pi^{\left(N_{n}+N_{\varepsilon}\right) / 2}} \sum_{N_{n}+N_{l}} \cdots \sum_{j=1}^{N_{n}+N_{\varepsilon}} \omega_{i_{j}} \\
& \cdot \mathbf{G}_{2,4}^{3,1}\left[\left.-\frac{2 \alpha \beta}{A_{0} \eta P_{t} \wp_{l}} \Phi\left(z_{i_{1}}, z_{i_{2}}, \cdots, z_{i_{N_{n}+N_{\varepsilon}}}\right) \right\rvert\, \begin{array}{l}
1,1+\rho^{2} \\
\alpha, \beta, \rho^{2}, 0
\end{array}\right] \tag{24}
\end{align*}
$$

where $N_{n}$ and $N_{\varepsilon}$ are the number of $n_{i}$ and $\varepsilon_{i}$, respectively. $\Phi\left(\omega_{i_{1}}, \omega_{i_{2}}, \cdots, \omega_{i_{n_{n}+n_{l}}}\right) \quad$ denotes the results of $\sum_{i=k}^{k+l} n_{i} \varsigma_{i}^{\mathbf{T}} \cdot \varepsilon_{i}$ with $\varsigma_{i}$ and $\varepsilon_{i}$ substituted by $z_{i_{1}}, z_{i_{2}}, \cdots, z_{i_{n_{n}+n_{\varepsilon}}}$.

Fig. 5 illustrates the conditional expectation $P\left(\Theta_{l} \mid h\right)$ with $l=3,4,5$. The simulation results are shown for different given

$$
\begin{align*}
P_{s}^{l o w} & =\frac{4}{\pi} \sum_{i} \sum_{j} \omega_{i} \omega_{j} \cdot Q\left(\frac{P_{t}^{2} \eta^{2} A^{2} h_{l}^{2} \rho^{2}(\alpha+1)(\beta+1)}{\alpha \beta\left(2+\rho^{2}\right) R_{b} \sigma_{n}^{2}+\left(z_{i}+z_{j}\right) P_{t}^{2} \eta^{2} A^{2} h_{l}^{2} \rho^{2}(\alpha+1)(\beta+1)}\right)  \tag{18}\\
P_{s}^{u p} & =4 \iiint Q\left(\tilde{\gamma}_{+-}\right) \cdot f_{h}(h) \cdot d h \cdot f_{\varsigma}\left(\varsigma_{k}\right) d \varsigma_{k} \cdot f_{\varsigma}\left(\varsigma_{k+1}\right) d \varsigma_{k+1} \\
& =\frac{16}{\pi} \sum_{i} \sum_{j} \omega_{i} \omega_{j} \cdot Q\left(\frac{P_{t}^{2} \eta^{2} A^{2} h_{l}^{2} \rho^{2}(\alpha+1)(\beta+1)}{\alpha \beta\left(2+\rho^{2}\right) R_{b} \sigma_{n}^{2}+\left(z_{i}+z_{j}\right) P_{t}^{2} \eta^{2} A^{2} h_{l}^{2} \rho^{2}(\alpha+1)(\beta+1)}\right)  \tag{19}\\
P_{s}^{\infty} & \approx \frac{1}{4} \iint\left[Q\left(\tilde{\gamma}_{++}^{\infty}\right)+Q\left(\tilde{\gamma}_{+-}^{\infty}\right)+Q\left(\tilde{\gamma}_{-+}^{\infty}\right)+Q\left(\tilde{\gamma}_{--}^{\infty}\right)\right] \cdot f_{\varsigma}\left(\varsigma_{k}\right) d \varsigma_{k} f_{\varsigma}\left(\varsigma_{k-1}\right) d \varsigma_{k-1} \\
& =\frac{1}{\pi} \cdot \sum_{i} \sum_{j} \omega_{i} \omega_{j}\left[2 Q\left(\frac{1-\sqrt{2} \sigma_{\xi} z_{i}}{\sqrt{2} \sigma_{\xi} z_{j}}\right)+Q\left(\frac{1}{\sqrt{2} \sigma_{\xi}\left(z_{i}+z_{j}\right)}\right)+Q\left(1-\sqrt{2} \sigma_{\xi}\left(z_{i}+z_{j}\right)\right)\right] \tag{20}
\end{align*}
$$

$h$, which are in accordance with theoretical ones. It's ensured that $P\left(\Theta_{3} \mid h\right)$ is much larger than $P\left(\Theta_{l} \mid h\right)$ with $l \geq 4$, i.e. $P\left(\Theta_{3} \mid h\right) \gg P\left(\Theta_{l} \mid h\right), l \geq 4$. As known to all, since the probability of continuous errors is much smaller than that of single error. In this sequel, $P\left(\Theta_{3} \mid h\right)$ could be the approximation value of $P\left(\Theta_{l} \mid h\right)$, which is $P\left(\Theta_{l} \mid h\right)>\approx P\left(\Theta_{3} \mid h\right)$. A similar approximation can be drawn by calculating the mathematical expectations of the conditional probabilities, which is $P\left(\Theta_{l}\right)>\approx P\left(\Theta_{3}\right)$. The expression of $P\left(\Theta_{3}\right)$ is furnished in Eq.(25).

$$
\begin{align*}
P\left(\Theta_{3}\right)= & \mathbb{E}_{h}\left[P\left(\Theta_{3} \mid h\right)\right] \\
= & \mathbb{E}_{h, n_{i}, \varsigma_{i}}\left[P\left(h<-2 / \eta P_{t} \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathbf{T}} \cdot \varepsilon_{i}\right)\right] \\
= & \frac{A_{0}}{2 \alpha \beta} \mathbb{E}_{h, n_{i}, \varsigma_{i}}\left[\mathbf { G } _ { 2 , 4 } ^ { 3 , 1 } \left(-\frac{2 \alpha \beta}{A_{0} \eta P_{t}}\right.\right. \\
& \left.\left.\cdot \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} \left\lvert\, \begin{array}{c}
1,1+\rho^{2} \\
\alpha, \beta, \rho^{2}, 0
\end{array}\right.\right)\right] \tag{25}
\end{align*}
$$

As stated in the Appendix A, there will be 16 cases of different signs of $\xi_{k-1}, \xi_{k}, \xi_{k+1}, \xi_{k+2}$. Therefore, the corresponding values of $\sum_{i=k}^{k+2} n_{i} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$ have 4 cases, which are $n_{k+1} \cdot\left(1-\varsigma_{k+1}\right)+n_{k+2} \cdot \varsigma_{k+1}, n_{k+1} \cdot\left(1-\varsigma_{k+1}\right)+n_{k}$. $\varsigma_{k+1}, n_{k+1}, n_{k} \cdot \varsigma_{k}+n_{k+1}\left(1-\varsigma_{k}-\varsigma_{k+1}\right)+n_{k+2} \cdot \varsigma_{k+1}$, respectively. Due to the fact that $n_{k}, n_{k+1}, n_{k+2}$ are independent and identically distributed, the first two cases has the same expectation. By adopting Gaussian-Hermite polynomials repeatedly, the closed form of $P\left(\Theta_{l}\right)$ is derived as

$$
\begin{align*}
P\left(\Theta_{l}\right) \approx & P\left(\Theta_{3}\right) \approx \frac{A_{0}}{\sqrt{\pi} \alpha \beta} \sum_{i} \omega_{i} \cdot \mathbf{G}_{2,4}^{3,1}\left(\frac{2 \sqrt{2} \sigma \alpha \beta}{A_{0} \eta P_{t}} z_{i}\right) \\
& +\frac{4 A_{0}}{\pi^{2} \alpha \beta} \cdot \sum_{i, j, k, p} \omega_{i} \omega_{j} \omega_{k} \omega_{p} \\
& \cdot \mathbf{G}_{2,4}^{3,1}\left(\frac{2 \alpha \beta}{A_{0} \eta P_{t}}\left(\sqrt{2} \sigma z_{i}-2 \sigma \sigma_{\xi}\left(z_{k} z_{p}-z_{i} z_{j}\right)\right)\right) \\
& +\frac{2 A_{0}}{\pi^{5 / 2} \alpha \beta} \sum_{i, j, k, p, q} \omega_{i} \omega_{j} \omega_{k} \omega_{p} \omega_{q} \\
& \cdot \mathbf{G}_{2,4}^{3,1}\left[\frac { 2 \alpha \beta } { A _ { 0 } \eta P _ { t } } \left(\sqrt{2} \sigma z_{k}\right.\right. \\
& \left.\left.+2 \sigma \sigma_{\xi}\left(z_{q} z_{p}+z_{i} z_{j}-z_{k} z_{i}-z_{k} z_{p}\right)\right)\right] \tag{26}
\end{align*}
$$

where $\mathbf{G}_{2,4}^{3,1}(\bullet)$ is short for $\mathbf{G}_{2,4}^{3,1}\left(\bullet \mid 1,1+\rho^{2} ; \quad \alpha, \beta, \rho^{2}, 0\right)$. Therefore, the BER results after TEVE could be derived as Eq.(27).

$$
\begin{equation*}
P_{s}^{T E V E}=\sum_{l=3}^{\infty} \wp_{l} \cdot P\left(\Theta_{l}\right) \approx P\left(\Theta_{3}\right) \tag{27}
\end{equation*}
$$

After deducing the BER expression of $P_{s}^{T E V E}$, let's make a comparison between the situations with or without the TEVE algorithm. Seeing that $P_{s}^{T E V E} \approx$ $\mathbb{E}_{h, n_{i}, \boldsymbol{\varsigma}_{i}}\left[P\left(h<-2 / \eta P_{t} \sum_{i=k}^{k+2} n_{i} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}\right)\right]$, we try to transform $P_{s}$ into $\mathbb{E}[P(h<\bullet)]$. Seeing that the BER $P_{s}$ (without TEVE) approaches its lower bound $P_{s}^{\text {low }}$ with high SNR, the task turns to transform $P_{s}^{\text {low }}$. According to Eq.(17), $P_{s}^{\text {low }}$ is equal to the expectation

TABLE II
Complexity Analysis of TEVE Algorithm

|  | Addition | Multiplication | Comparator |
| :--- | :---: | :---: | :---: |
| $\lambda_{k}^{++}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ | $3 / 2$ | $9 / 2$ | - |
| $\lambda_{k}^{+-}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ | $23 / 8$ | 9 | - |
| $\lambda_{k}^{-+}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ | $1 / 2$ | 1 | - |
| $\lambda_{k}^{--}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ | $3 / 2$ | $9 / 2$ | - |
| $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ | $51 / 8$ | 19 | - |
| TEVE algorithm | $51 \times 2^{\delta-3} \cdot \delta$ | $19 \times 2^{\delta} \cdot \delta$ | $3 \times 2^{\delta}-1$ |

$\mathbb{E}_{h, \varsigma_{k}, \varsigma_{k+1}}\left[Q\left(\frac{P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\left(\varsigma_{k}+\varsigma_{k+1}\right) P_{t}^{2} h^{2} \eta^{2}}\right)\right] . \quad$ By $\quad$ introducing the variable $\vartheta$ obeying the standard normal distribution (i.e. $\vartheta \sim N(0,1)$ ), the expectation of $P_{s}^{\text {low }}$ becomes $\mathbb{E}_{\vartheta, h, \varsigma_{k}, \varsigma_{k+1}}\left[\vartheta>\left(\frac{P_{t}^{2} h^{2} \eta^{2}}{R_{b} \sigma_{n}^{2}+\left(\varsigma_{k}+\varsigma_{k+1}\right) P_{t}^{2} h^{2} \eta^{2}}\right)\right]$. By the shift terms and adjustment of coefficients, this expectation is deduced as

$$
\begin{equation*}
P_{s}^{\text {low }}=\mathbb{E}_{\vartheta, h, \varsigma_{k}, \varsigma_{k+1}}\left[h<\frac{1}{P_{t} \eta} \sqrt{\left|\frac{\vartheta R_{b} \sigma_{n}^{2}}{1-\vartheta\left(\varsigma_{k}+\varsigma_{k}\right)}\right|}\right] \tag{28}
\end{equation*}
$$

By comparing Eq.(25) (BER with TEVE) and Eq.(28) (BER without TEVE), they both have the forms of $\mathbb{E}[P(h<\bullet)]$. Besides, they also share the same coefficient $\frac{1}{P_{t} \eta}$, which can be ignored. In this sequel, we can compare the expectation between $-2 \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathrm{T}} \cdot \varepsilon_{i}$ and $\sqrt{\left|\frac{\vartheta R_{b} \sigma_{n}^{2}}{1-\vartheta\left(\varsigma_{k}+\varsigma_{k}\right)}\right|}$. For brevity, we omit the expectation symbol $\mathbb{E}$. There are two main reasons that the former is much smaller than the latter, i.e. $P_{s}^{T E V E} \ll$ $P_{s}^{l o w}$. The first reason is that $-2 \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathrm{T}} \cdot \varepsilon_{i}$ may be negative, because the coefficients $n_{k}, n_{k+1}, n_{k+2}$ can be positive or negative while $\left.\boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}\right|_{i=k, k+1, k+2}$ keeps positive. The circumstance of negative $-2 \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathrm{T}} \cdot \varepsilon_{i}$ makes no contribution to $P_{s}^{T E V E}$, since $h$ is a positive variable. The second reason is that the positive $-2 \sum_{i=k}^{k+2} n_{i} \varsigma_{i}^{\mathrm{T}} \cdot \varepsilon_{i}$ has a great probability to be smaller than $\sqrt{\left|\frac{\vartheta R_{b} \sigma_{n}^{2}}{1-\vartheta\left(\varsigma_{k}+\varsigma_{k}\right)}\right|}$. We will make a comparison by order of magnitude. The coefficients $n_{k}, n_{k+1}, n_{k+2}$ have the same order of its standard deviation $\sqrt{R_{b}} \sigma_{n}$. Therefore, the former's magnitude is dominated by $-2 \sum_{i=k}^{k+2} \varsigma_{i}^{\mathrm{T}} \cdot \varepsilon_{i}$. The latter's magnitude is dominated by $\sqrt{\left|\frac{\vartheta}{1-\vartheta\left(\varsigma_{k}+\varsigma_{k}\right)}\right|}$, whose magnitude can be approximated by deleting the denominator less than 1, i.e. $\sqrt{|\vartheta|} \approx 0.82$. It's apparent that the former's magnitude is smaller than $\sqrt{|\vartheta|}$. It indicates that our TEVE algorithm can mitigate the impairment by timing errors, which will be further validated by experimental results in Sec.V.

Complexity of equalization algorithm is critical for practical applications. Thus we furnish the computation complexity analysis. As shown in Eq.(14), four kinds of BMs $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ could be calculated with the variables of $x_{k-1}, x_{k}, x_{k+1}$ and $\mu_{k}$. It seems that we have to calculate all the squared items and cross items. Noting that the addition and multiplication will not be counted, when one or more parameters participating in computing are equal to zero. It would reduce the complexity, when the vector $\mathbf{x}_{\mathbf{k}}$ has more zero elements. Table II illustrates the mean complexity of the TEVE algorithm in each case. It's not strange that number of adding operation is not an integer, because it shows the mean


Fig. 6. An example of an independent "bad event" and an example of a "bad event" returning to a "clean event".
value with different $\mathbf{x}_{k}=[0$ or 1,0 or 1,0 or 1$]$ and different $\Omega_{k}= \pm \pm$.

## C. Analysis on "Bad Events"

In this subsection, the "bad events" will be illustrated and analyzed. As defined in Sec.II, the "bad event" means the timing error's absolute value $\varsigma$ is larger than 1 . We may make the assumption that $\xi_{k+1}$ is larger than 1. ${ }^{2}$ Fig. 6 depicts the corresponding situation of "bad events." Although a "bad event" has happened in the $k+1$-th bit, it will usually come back to the "clean event" from the $(k+2)$-th bit, i.e. $\varsigma_{k+2}=\left|\xi_{k+2}\right|<1$. This is the normal circumstance, because the probability of two or more adjacent "bad events" can be approximated to zero. For instance, the probability of two consecutive "bad events" is equal to $3.28 \times 10^{-13}$ with $\sigma_{\xi}=0.2$. And the probability turns to $4.88 \times 10^{-32}$ when $\sigma_{\xi}$ is equal to 0.1 . In this section, we furnish the case of an independent "bad event". The situations of two or more continuous "bad events" can be also analyzed in the similar way.

In an arbitrary independent "bad event", two or more neighboring bits participate in the interference items. There will be three situations with respect to different $\xi_{k+2}$, which are $\xi_{k+2}<0,0<\xi_{k+2}<\xi_{k+1}-1$ and $\xi_{k+2}>\xi_{k+1}-1$, respectively. ${ }^{3}$ We need to mention that the adjacent edges in the order of appearance constitute the clock during the clock generation progress. Therefore, in the first case, there is one edge in either the $(k+1)$-th or $k$-th bit interval. In other words, the first case returns to the "clean event", as shown in the third row of Fig.6. Moreover, the latter two cases can be summarized as the same situation. In the latter two cases, there is no positive edge in the $k$-th bit interval and there are two positive edges in the $(k+2)$-th bit interval. Thus we may consider the case of $\xi_{k+2}>\xi_{k+1}-1$ as an example, which is shown in the second row of Fig.6. In this sequel,

[^1]

Fig. 7. Simulated results of "bad events" with the TEVE algorithm (solid lines with solid markers) or without the TEVE algorithm (dash-dotted lines with hollow markers).
$\mu_{k}, \mu_{k+1}, \mu_{k+2}$ can be expressed in Eq.(29).

$$
\begin{align*}
\mu_{k}= & \eta P_{t} h\left[\varsigma_{k} \cdot x_{k-1}+x_{k}+x_{k+1}\right. \\
& \left.+\left(\varsigma_{k+1}-1\right) \cdot x_{k+2}\right]+n_{k} \\
\mu_{k+1}= & \eta P_{t} h\left[\left(\varsigma_{k+2}-\varsigma_{k+1}+1\right) \cdot x_{k+2}\right]+n_{k+1} \\
\mu_{k+2}= & \eta P_{t} h\left[\left(1-\varsigma_{k+2}\right) \cdot x_{k+2}+\varsigma_{k+3} \cdot x_{k+3}\right]+n_{k+2} \tag{29}
\end{align*}
$$

where $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right) \Xi\left(\xi_{k+2}\right) \Xi\left(\xi_{k+3}\right) \quad$ is assumed to be +--- .

From Eq.(29), $\mu_{k}$ has three interference items ( $x_{k-1}, x_{k+1}$ and $x_{k+2}$ ). Compared with Eq.(5) of "clean events", more interference items are added for the $k$-th bit. Moreover, the signal item should be $x_{k+1}$ in $\mu_{k+1}$. However, the expression of $\mu_{k+1}$ has nothing to do with $x_{k+1}$ due to the "bad event" in $k+1$-th bit, which may to a bit error. From the expression of $\mu_{k+2}$, we can notice that it will turn back to the "clean event" since $(k+2)$-th bit. That is to say, the "bad event" will spread no negative effects for subsequent bits.

In order to obtain the performance of our TEVE algorithm in "bad events", the expressions of $\mu_{k}, \mu_{k+1}, \mu_{k+2}$ can be substituted to Eq.(14) in order to obtain the corresponding BM. In this sequel, Fig. 7 illustrates the simulation results of the TEVE algorithm in the "bad events" with solid lines and solid markers, while the results without TEVE algorithm are also given in dash-dotted lines with hollow markers. As seen from Fig.7, the damage of the "bad events" to system performance can be slightly compensated. The reasons are given as follows. When a "bad event" happens, it has the probability of returning to a "clean event," such as the case of $\xi_{k+1}>1$ and $\xi_{k+2}<0$ in the third row of Fig.6. In this case, the timing errors in "clean events" can be mitigated by our TEVE algorithm. In a more general case, let's discuss the situations that "bad events" cannot return to "clean events". For an independent "bad event" on $\xi_{k+1}$, the decision progress of $x_{k}$ can still benefit from our TEVE algorithm. Considering $\mu_{k}$ in Eq.(29), the BMs $\lambda_{k}^{ \pm \pm}\left(\mu_{k} \mid \mathbf{x}_{k}\right)$ calculate expected squared distances between $\mu_{k}$ and


Fig. 8. The scene of indoor experiment (left) and the structural block diagram (Right).


Fig. 9. Eye diagram of signals impaired by timing errors (a) $\sigma_{\xi}=0.05$, (b) $\sigma_{\xi}=0.1$ and (c) $\sigma_{\xi}=0.2$.
$\eta P_{t} h \mathbf{x}_{k}^{\mathbf{T}} \cdot \boldsymbol{\varsigma}_{k}^{ \pm \pm}$. Among four possible signs $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=$ $\pm \pm$, it's apparent that $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=+-$ case has the smallest distance than the other three cases, where $\lambda_{k}^{+-}\left(\mu_{k} \mid \mathbf{x}_{k}\right)=\mathbb{E}\left\{\left[\mu_{k}-\eta P_{t} h\left(\varsigma_{k} x_{k-1}+x_{k}+\varsigma_{k+1}\right.\right.\right.$ $\left.\left.\left.x_{k+1}\right)\right]^{2}\right\}$. In this sequel, the TEVE algorithm may make correct decision on $x_{k}$ with a certain large probability, even though a "bad event" happens on $\xi_{k+1}$. It needs to mention that the conditional BERs $P_{s \mid b a d}$ in the magnitudes of $10^{-2}$ does not dominate the system's performance, since the probabilities $P_{b a d}$ of "bad events" are rather low. That is to say, the average BER in "bad events" makes a ignorable contribution to the total average BER result, as analyzed in Eq.(3).

## V. Experiments and Theoretical Performance

After deducing the closed forms of $P_{s}$ and $P_{s}^{T E V E}$, the indoor experiments are elaborated in this section. Both the theoretical performances and experimental ones are compared with each other. Fig. 8 shows the structural block diagram, as well as the scene of the experimental system. During the experiment, the signals are created by the Arbitrary Wave Generator (AWG, Tektronix 70002A). Thanks to the software "SerialXpress" embedded in the AWG, we have the ability of producing the signals suffered from various timing errors [43]. By utilizing the real-time oscilloscope (Tektronix DPO73304D, 33GHz, 100Gsps), we can validate the ISI issue brought by timing errors. Fig. 9 illustrates these eye diagrams of the AWG's outputs with $\sigma_{\xi}=0.05,0.1,0.2$, respectively. It's noted that the situation that we sample the ideal signal with the clock suffered from timing errors can be approximately equivalent to the situation that we sample the signals impaired by timing errors with ideal clock. The two
situations have nearly the same eye diagrams. In these eye diagrams, the horizontal axis is divided into 8 units of 100 ps , while the range of the vertical axis is 600 mV with a uniform unit to be 100 mV . Every symbol should have 4 units in the ideal case, when data rate $R_{b}$ is equal to 2.5 Gbps . As can be seen from Fig.9, the eye diagrams get worse with the increase of $\sigma_{\xi}$. In the case of $\sigma_{\xi}=0.05$, the eye keeps open with a few messy edges in the middle region and blurred borders around the cross points. In the case of $\sigma_{\xi}=0.1$, the eye diagram can hardly be opened, while we can barely see any shapes of eyes in the worst case of $\sigma_{\xi}=0.2$.

After modulated by the Mach-Zehnder modulator (MZM), optical modulation signal enters the space with the help of the transmitting lens. ${ }^{4}$ The receiver couples the collected signal into a multimode fiber. Then the turbulence channel is simulated by the atmosphere scintillation playback system (ACPS), where the data of channel gain would be transmitted from the host computer by the Ethernet interface. ${ }^{5}$ The ACPS has two fiber interfaces as the input and output. It can attenuate the input optical signal dynamically ranging from 0 dB to 30 dB , with the minimum time resolution of $1 \mu \mathrm{~s}$. According to the statistical distribution in Eq.(1), the variables of channel gains can be generated and further transferred into the ACPS. In order to verify the channel fluctuation, we employ an optical power meter with the sampling rate of 10 ksps to collect the optical power out of the ACPS. The logarithmic amplifier is

[^2]

Fig. 10. The simulated fading channel achieved by the ACPS (a) Channel Fluctuation (b) Power Statistics.
utilized in the optical power meter to obtain a large dynamic range. Fig.10(a) and 10(b) depict the instantaneous optical power and the corresponding histogram, respectively. As can be seen from Fig.10(a), the range of fluctuation is about 30 dB . In Fig.10(b), the bars denote the histogram, while the red line stands for the theoretical PDF in Eq.(1). Due to the fact that the histogram (also shown as the pseudo-color parts in Fig.10(a)) confirms the theoretical PDF, we can draw the conclusion that the fading channel caused by turbulence and pointing errors can be simulated accurately by the ACPS during the experiment.

As expounded in Fig.8, there will be two methods of measuring the received signal. One is directly measurement of the received signal by the optical sampling oscilloscope (Tektronix DSA8300), while the other way is to sample the electric signal by A/D converter, after photoelectric conversion and amplified by trans-impedance amplifier (TIA). Then, we store the sampled data by FPGA (Xilinx xc7k325t), and transmit them to host PC through Ethernet. The stored data can be also analyzed in an offline way by the host PC.
The software 80 JNS embedded in DSA8300 has the ability of measuring the features of optical signals. Different from the real-time oscilloscope, the optical sampling oscilloscope DSA8300 needs the clocking input, which is homologous with the signals to be measured. The clock can be provided by the second channel output of the AWG. Fig. 11 depicts the measurement results of optical signals coupled from the receiving lens by DSA8300. The yellow bars in Fig.11(a) illustrate the histogram of timing errors (i.e. total jitters, TJ),

(a) Histogram of Total Jitter

(b) BER Eye Diagram

Fig. 11. DSA 8300's results of the received optical signals impaired by timing errors and fading channels.
which almost exactly fits the red line (the PDF of timing errors $\xi$ with $\sigma_{\xi}=0.04$ ). The perfect fitting curves of both the timing errors in Fig.11(a) and fading channels in Fig. 10 ensure that the experimental results are consistent with the theoretical results. Fig.11(b) denotes the BER eye diagram. The eyes can barely open in the vertical and horizontal directions. The eye height and eye width are influenced by the channel fading and timing errors, respectively. Their vertical axis shows that the optical power ranges from 0 to $70 \mu \mathrm{~W}$ (i.e. less than -11.55 dBm ) with most curves lying below $20 \mu \mathrm{~W}$ $(-17 \mathrm{dBm})$, which indicates the channel fluctuation described in Fig. 10 indirectly. Note that the maximum value of the vertical axis (i.e. -11.55 dBm ) is the greatest value of optical power during a time window of 32768 symbols (i.e.about $13 \mu \mathrm{~s}$ ) by DSA8300, which is not contrary to the maximum value of -10.33 dBm in Fig.10. According to the Fig.11(a), the BER cannot approach 0 infinitely, even with the optimal decision point. It mainly results from the ISI caused by timing errors, as given in Fig.9(a). From Fig. 9 and Fig.11, it's explicit that the timing error $\xi$ cripples the system performance.
Fig. 12 compares the BER results of TEVE algorithm and the original performance. The lines denote the theoretical results, which are nearly in agreement with the markers representing the experimental results. The parameters are shown in Table III. The coincident results are achieved thanks to the accurate channel simulation and production of timing errors. It could be found from Fig. 12 that the BER decreases with larger standard deviation $\sigma_{\xi}$. However, the slope increases first and then diminishes. It can be explained that the ISI caused by smaller $\sigma_{\xi}$ could be ignored, at that time, the BER is mainly impaired by the channel gain $h$. The initial points are about $6 \times 10^{-7}$ for $\alpha=4, \beta=2$ and $5 \times 10^{-6}$ with $\alpha=2, \beta=1$.


Fig. 12. The BER results with or without the TEVE algorithm and the adaptive DFE method versus different $\sigma_{\xi}$ by both experiments and theoretical lines or bounds.

TABLE III
Parameters in Monte-Carlo Simulations

| Parameters | Value |
| :--- | :--- |
| Responsivity $\eta$ | $0.82 A / W @ 1550 \mathrm{~nm}$ |
| Propagation Distance $z$ | 20 km |
| Receiving Aperture $a$ | 0.1 m |
| Beam Radius $w_{z}$ at 1 km | 2.5 m |
| Misalignment Jitter Standard Deviation $\sigma_{s}$ | 0.2 m |
| Effective Number of Eddies $(\alpha, \beta)$ | $(4,2)$ or $(2,1)$ |
| Wavelength $\lambda$ | 1550 nm |
| Data Rate $R_{b}$ | 2.5 Gbps |

The two values can be considered as the ideal situation where the system does not suffer from timing errors. The timing error $\xi$ hampers the BER so much with $\sigma_{\xi}>0.07$ that almost has no difference with much larger $\sigma_{\xi}$. Additionally, the results without TEVE algorithm are lied between the upper and lower bounds deduced in Sec.IV-A. The lower bound is always equal to a quarter of the upper bound as the expressions in Eq.(17) and Eq.(19).

In order to verify the effectiveness of our TEVE algorithm, the results of adaptive DFE are also given as the control group. For a fair comparison, four different DFE equalizers are designed to work in four cases of $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)= \pm \pm$. The furnished results with traditional adaptive DFE (marked as "*") takes the mean value of the four equalizers' BER results. It could be maintained from Fig. 12 that our TEVE algorithm outperforms the adaptive DFE method. The main reason is that the ISI caused by timing errors varies so quickly that the hysteresis feature of DFE method can bring about bit errors. Besides, the wrong judgments of the previous symbols can lead to the error propagation. That's why the DFE performs even worse than the situation without equalization, when the $\sigma_{\xi}$ is large enough, i.e. $\sigma_{\xi}>0.07$ with $\alpha=2, \beta=1$. It's also evaluated that the TEVE algorithm could mitigate BER introduced by the timing error $\xi$ as much as three orders of magnitude.

Fig. 13 represents the results with high SNR. In order to achieve the experimental results with high SNR, the avalanche


Fig. 13. Comparison of the BER performance with or without the TEVE algorithm in the case of high power for $\sigma_{\xi}=0.05,0.1,0.2$.
photodiode (APD) should be replace by the positive-intrinsicnegative (PIN) photodiode. The latter can tolerate high power at the sacrifice of sensitivity. The maximum value of horizontal axis is limited the bearable optical power of PIN. In Fig.13, the slopes of solid lines without TEVE algorithm cannot approach 0 , which indicates the asymptotic bounds analyzed in Sec.IV-A. The gap between the solid lines and dash-dotted lines becomes larger. That is to say, the TEVE algorithm has better performance with the increasing power. It's also concluded from Fig. 12 and Fig. 13 that our TEVE algorithm cannot eliminate the BER degradation, as analyzed in Sec.IV-B. In addition, the experimental results perform a little worse than the theoretical results whether we employ the TEVE algorithm or not. In the case without TEVE algorithm, the theoretical values are calculated under "clean events", whose probability is slightly smaller than the (i.e. $P_{s \mid c l}<\approx P_{s}$ ). This reason also can be applied to the case with TEVE algorithm. Another reason of TEVE algorithm is by utilizing the approximation $P\left(\Theta_{3}\right) \approx<P\left(\Theta_{l}\right)$ in the derivation of the theoretical results.

## VI. CONCLUSION

Timing errors bring about the ISI to the OWC systems, leading to an increase on the BER. Based on MLSD, the TEVE algorithm has been proposed to mitigate the injury. Considering the two possible signs $\Xi\left(\xi_{k}\right)= \pm$ for each symbol, the number of states has doubled compared to the traditional Viterbi algorithm. The closed forms of BMs between two continuous states are deduced in four cases of different $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)= \pm \pm$ by the mathematical expectation of $h, \varsigma_{k}, \varsigma_{k+1}$. Each CM is equal to the summation of BMs. By reserving the path with largest CM and discarding others, the symbols are detected. Then we compare the BER results with or without the TEVE algorithm. The BER performance of timing errors is analyzed for four different situations of $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)= \pm \pm$. It's obtained that $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=$ +- is the worst case, while $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right)=-+$ is the most moderate case. We deduce the upper bound $P_{s}^{u p}$, lower bound $P_{s}^{l o w}$ by mathematical enlarging and reducing method. What's more, the asymptotic boundary $P_{s}^{\infty}$ is also elaborated, which indicates the error floor caused by timing errors. We also

TABLE A1
The Results of $\sum_{i=k}^{k+l-1} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$ With $l=3$ By Enumeration Method

| $\Xi\left(\xi_{k} \sim \xi_{k+3}\right)$ | $\varsigma_{k}^{\mathbf{T}} \cdot \varepsilon_{k}$ | $\varsigma_{k+1}^{\mathbf{T}} \cdot \varepsilon_{k+1}$ | $\varsigma_{k+2}^{\mathbf{T}} \cdot \varepsilon_{k+2}$ | $\sum_{i=k}^{k+2} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$ |
| :---: | :---: | :---: | :---: | :---: |
| + + ++ | 0 | $1-\boldsymbol{S}_{k+2}$ | $S_{k+2}$ | 1 |
| + + + | 0 | $1-s_{k+2}$ | $\varsigma_{k+2}$ | 1 |
| $++-+$ | 0 | 1 | 0 | 1 |
| + + -- | 0 | 1 | 0 | 1 |
| + - ++ | $\boldsymbol{S}_{k+1}$ | $1-\boldsymbol{s}_{\boldsymbol{k}+1}-\boldsymbol{s}_{\boldsymbol{k}+2}$ | $\varsigma_{k+2}$ | 1 |
| + - +- | $\boldsymbol{S}_{k+1}$ | $1-\boldsymbol{s}_{\boldsymbol{k}+1}-\boldsymbol{s}_{\boldsymbol{k}+2}$ | $s_{k+2}$ | 1 |
| + - -+ | $S_{k+1}$ | $1-\boldsymbol{S}_{k+1}$ | 0 | 1 |
| + - -- | $\varsigma_{k+1}$ | $1-\boldsymbol{s}_{\boldsymbol{k}+1}$ | 0 | 1 |
| - + ++ | 0 | $1-\boldsymbol{S}_{\boldsymbol{k}+2}$ | $S_{k+2}$ | 1 |
| - + + - | 0 | $1-\boldsymbol{s}_{k+2}$ | $\varsigma_{k+2}$ | 1 |
| - + -+ | 0 | 1 | 0 | 1 |
| - + -- | 0 | 1 | 0 | 1 |
| - - ++ | $s_{k+1}$ | $1-\boldsymbol{s}_{k+1}-\boldsymbol{s}_{k+2}$ | $\varsigma_{k+2}$ | 1 |
| --+- | $S_{k+1}$ | $1-\boldsymbol{S}_{k+1}-\boldsymbol{S}_{k+2}$ | $S_{k+2}$ | 1 |
| ---+ | $\varsigma_{k+1}$ | $1-\boldsymbol{S}_{k+1}$ | 0 | 1 |
| ---- | $\varsigma_{k+1}$ | $1-\boldsymbol{S}_{k+1}$ | 0 | 1 |

furnish the theoretical approximation expressions of BER after TEVE algorithm, which matches the experimental results tightly. It's evaluated that the TEVE algorithm is able to mitigate the ISI-introduced BER as much as three orders of magnitude. Besides, the TEVE algorithm can still provide nearly two magnitude enhancement on BER with high SNR even in the terrible case of $\sigma_{\xi}=0.2 \mathrm{UI}$. In addition, it's possible for the TEVE algorithm to be employed in the practical system, in the light of the complexity analysis results. This work hopes to make a contribution to achieving reliable transmission in the OWC systems with higher rates. The combination of our TEVE algorithm and Channel Coding will be further analyzed in our future research.

## Appendix A

By adopting mathematical induction, LEMMA 1 is proved. First, when $l=3$, there would be 16 kinds of $\sum_{i=k}^{k+2} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$, due to different $\Xi\left(\xi_{k}\right) \Xi\left(\xi_{k+1}\right) \Xi\left(\xi_{k+2}\right) \Xi\left(\xi_{k+3}\right)$. The results are enumerated in Table A1 1. It could be conclude that $\sum_{i=k}^{k+2} \varsigma_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$ is proved to be $\wp_{3}$.

Then we assume the hypothesis of $\sum_{i=k}^{k+l_{0}-1} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}=\wp_{l_{0}}$ is valid for any arbitrary integer $l$ satisfying $l \leq l_{0},\left(l_{0} \geq 3\right)$. All we need to do is to prove the equation $\sum_{i=k}^{k+l-l_{0}} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}=\wp_{l_{0}+1}$. The summation could be expanded as

$$
\begin{equation*}
\sum_{i=k}^{k+l_{0}} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}=\boldsymbol{\varsigma}_{k}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{k}+\sum_{i=k+1}^{k+l_{0}} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} \tag{A1}
\end{equation*}
$$

Obviously, the vectors $\left[\varepsilon_{k+2}, \ldots, \varepsilon_{k+l_{0}}, \varepsilon_{k+l_{0}+1}\right]$, $\left[\varepsilon_{k+3}, \ldots, \varepsilon_{k+l_{0}}, \varepsilon_{k+l_{0}+1}\right], \ldots,\left[\varepsilon_{k+l_{0}-1}, \varepsilon_{k+l_{0}}, \varepsilon_{k+l_{0}+1}\right]$ ending with $[1,0,0]$ and starting with 1 could be considered as the subset of $\varepsilon_{l_{0}+1}^{e r}=\left[\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \boldsymbol{\varepsilon}_{k+l_{0}}, \boldsymbol{\varepsilon}_{k+l_{0}+1}\right]$. According to REMARK 1, the values of these vectors are equal to $\varepsilon_{l_{0}}^{e r}, \varepsilon_{l_{0}-1}^{e r}, \ldots, \varepsilon_{3}^{e r}$, respectively. However, the summation $\sum_{i=k+1}^{k+l_{0}} \varsigma_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}$ could be further simplified as Eq. (A2).

$$
\begin{aligned}
& \sum_{i=k+1}^{k+l_{0}} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} \\
& =[0,1,0] \cdot \varepsilon_{k+1}+[0,0,1] \cdot \boldsymbol{\varepsilon}_{k+2}
\end{aligned}
$$

TABLE A2
Main Variables and Their Brief Definitions

| Name | Paraphrase |
| :--- | :--- |
| $\alpha$ | effective number of large eddies |
| $\beta$ | effective number of small eddies |
| $\eta$ | responsivity of the photodiode |
| $\Lambda_{ \pm_{ \pm}}^{k_{ \pm}}$ | cumulative metric for $k$-th bit |
| $\lambda_{k}$ | branch metric for $k$-th bit |
| $\Omega_{k}$ | $\Xi\left(\xi_{k}\right)$ and $\Xi\left(\xi_{k+1}\right)$ |
| $\omega_{i}$ | root of $i$-th-order Hermite polynomials |
| $\rho$ | ratio between the equivalent beam radius and $\sigma_{s}$ |
| $\sigma_{\xi}$ | standard deviation of $\xi_{k}$ |
| $\sigma_{s}$ | standard deviation of misalignments |
| $\Theta_{l}$ | error event with the length of $l$ |
| $\tilde{\gamma}_{ \pm \pm}$ | SINR for different $\Omega_{k}$ |
| $\tilde{\gamma}_{ \pm \pm}^{ \pm}$ | asymptotic value of SINR $\tilde{\gamma}_{ \pm \pm}$ |
| $\mu_{k}$ | $k$-th electric signal in the receiver |
| $\varepsilon_{i}$ | vector $\left[\varepsilon_{i-1}, \varepsilon_{i}, \varepsilon_{i+1}\right]$ |
| $\varsigma_{k}$ | vector $\left[\varsigma_{k-1}, \varsigma_{k}, \varsigma_{k+1}\right]$ |
| $\varepsilon_{l}^{e r}$ | error vector of the error event $\Theta_{l}$ |
| $\varepsilon_{i}$ | error difference between $x_{i}$ and $\hat{x}_{i}$ |
| $\varsigma_{k}$ | absolute value of $\xi_{k}$ |
| $\wp_{l}$ | number of "1" in $\varepsilon_{l}^{e r}$ |
| $\Xi\left(\xi_{k}\right)$ | sign of $\xi_{k}$ |
| $\xi_{k}$ | timing error for the $k$-th bit |
| $\mathbf{x}_{k}$ | vector $\left[x_{k-1}, x_{k}, x_{k+1}\right]$ |
| $\hat{x}_{k}$ | $k$-th detected data |
| $A_{0}$ | maximum fraction of the collected power |
|  | receiving lens |
| $h$ | channel gain |
| $h_{l}$ | path attenuation |
| $N_{\varepsilon}$ | number of $\varepsilon_{i}$ |
| $N_{n}$ | number of $n_{i}$ |
| $P_{t}$ | transmitted power |
| $R_{b}$ | data rate |
| $T_{b}$ | bit interval |
| $x_{k}$ | $k$-th transmitted data |
| $z_{i}$ | weight of $i$-th-order Hermite polynomials |
|  |  |

$$
\begin{align*}
& +\left.\sum_{i=k+1}^{k+l_{0}} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i}\right|_{\varepsilon_{k+1}=0} \\
= & {[0,1,0] \cdot \boldsymbol{\varepsilon}_{k+1}+[0,0,1] \cdot \boldsymbol{\varepsilon}_{k+2}+\sum_{i=k}^{k+l_{0}-1} \boldsymbol{\varsigma}_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} } \\
= & {[0,1,0] \cdot \boldsymbol{\varepsilon}_{k+1}+[0,0,1] \cdot \boldsymbol{\varepsilon}_{k+2}+\wp_{l_{0}} } \tag{A2}
\end{align*}
$$

Substituting Eq.(A2) to Eq.(A1), we can make the simplification in Eq. (A3).

$$
\begin{align*}
& \sum_{i=k}^{k+l_{0}} \varsigma_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} \\
& \quad=\varsigma_{k}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{k}+\sum_{i=k+1}^{k+l_{0}} \varsigma_{i}^{\mathbf{T}} \cdot \boldsymbol{\varepsilon}_{i} \\
& =[1,0,0] \cdot \boldsymbol{\varepsilon}_{k}+[0,1,0] \cdot \boldsymbol{\varepsilon}_{k+1}+[0,0,1] \cdot \boldsymbol{\varepsilon}_{k+2}+\wp_{l_{0}} \\
& =\wp_{3}+\wp_{l_{0}}=\wp_{l_{0}+1} \tag{A3}
\end{align*}
$$

At this point, the proof of LEMMA 1 is finished.

## Appendix B

The main variables and their definitions are illustrated in Table A2.

## REFERENCES

[1] A. Belmonte, M. T. Taylor, L. Hollberg, and J. M. Kahn, "Effect of atmospheric anisoplanatism on earth-to-satellite time transfer over laser communication links," Opt. Exp., vol. 25, no. 14, pp. 15676-15686, 2017.
[2] A. Al-Kinani, C.-X. Wang, L. Zhou, and W. Zhang, "Optical wireless communication channel measurements and models," IEEE Commun. Surveys Tuts., vol. 20, no. 3, pp. 1939-1962, 3rd Quart., 2018.
[3] H.-L. Zhou et al., "Orbital angular momentum complex spectrum analyzer for vortex light based on the rotational Doppler effect," Light Sci. Appl., vol. 6, no. 16251, pp. 1-8, 2016.
[4] D.-N. Nguyen, J. Bohata, M. Komanec, S. Zvanovec, B. Ortega, and Z. Ghassemlooy, "Seamless 25 GHz transmission of LTE 4/16/64-QAM signals over hybrid SMF/FSO and wireless link," J. Lightw. Technol., vol. 37, no. 24, pp. 6040-6047, Dec. 15, 2019.
[5] J. Li, M. Zhang, D. Wang, S. Wu, and Y. Zhan, "Joint atmospheric turbulence detection and adaptive demodulation technique using the CNN for the OAM-FSO communication," Opt. Exp., vol. 26, no. 8, pp. 10494-10508, 2018.
[6] M. Rahaim and T. D. C. Little, "Interference in IM/DD optical wireless communication networks," J. Opt. Commun. Netw., vol. 9, no. 9, p. D51, 2017.
[7] J. Zhao and L.-K. Chen, "Adaptively loaded IM/DD optical OFDM based on set-partitioned QAM formats," Opt. Exp., vol. 25, no. 8, pp. 9368-9377, 2017.
[8] B. Robinson, D. Boroson, D. A. Burianek, and D. V. Murphy, "The lunar laser communications demonstration," in Proc. IEEE Int. Conf. Space Opt. Syst. Appl. (ICSOS), Santa Monica, CA, USA, 2011, pp. 54-57.
[9] Z. Sodnik et al., "LLCD operations using the Lunar Lasercom OGS terminal," Proc. SPIE, vol. 8971, pp. 1-13, Mar. 2014.
[10] M. M. Willis et al., "Downlink synchronization for the lunar laser communications demonstration," in Proc. Int. Conf. Space Opt. Syst. Appl. (ICSOS), Santa Monica, CA, USA, May 2011, pp. 83-87.
[11] E. A. Park, D. Cornwell, and D. Israel, "NASA's next generation $\geq 100$ Gbps optical communications relay," in Proc. IEEE Aerosp. Conf., Big Sky, MT, USA, Mar. 2019, pp. 1-9.
[12] D. J. Israel, B. L. Edwards, and J. W. Staren, "Laser communications relay demonstration (LCRD) update and the path towards optical relay operations," in Proc. IEEE Aerosp. Conf., Big Sky, MT, USA, Mar. 2017, pp. 1-6.
[13] M. Chen, C. Liu, D. Rui, and H. Xian, "Performance verification of adaptive optics for satellite-to-ground coherent optical communications at large zenith angle," Opt. Exp., vol. 26, no. 4, pp. 4230-4242, 2018.
[14] V. W. S. Chan, "Free-space optical communications," IEEE/OSA J. Lightw. Technol., vol. 24, no. 12, pp. 4750-4762, Dec. 2006.
[15] X. Zhu and J. M. Kahn, "Free-space optical communication through atmospheric turbulence channels," IEEE Trans. Commun., vol. 50, no. 8, pp. 1293-1300, Aug. 2002.
[16] L. C. Andrews and R. L. Phillips, Laser Beam Propagation Through Random Media, 2nd ed. Bellingham, WA, USA: SPIE, 2005.
[17] Y. Wang et al., "Non-data-aided cycle slip self-correcting carrier phase estimation for QPSK modulation format of coherent wireless optical communication system," IEEE Access, vol. 7, pp. 110451-110462, 2019.
[18] J. Fang et al., "Polar-coded MIMO FSO communication system over gamma-gamma turbulence channel with spatially correlated fading," IEEE/OSA J. Opt. Commun. Netw., vol. 10, no. 11, pp. 915-923, Nov. 2018.
[19] I. S. Ansari, F. Yilmaz, and M.-S. Alouini, "Performance analysis of free-space optical links over Málaga $(\mathcal{M})$ turbulence channels with pointing errors," IEEE Trans. Wireless Commun., vol. 15, no. 1, pp. 91-102, Jan. 2016.
[20] Y. Wang and H. Cao, "An upper bound on the capacity of distributed MIMO systems with rician/lognormal fading," IEEE Commun. Lett., vol. 20, no. 2, pp. 312-315, Feb. 2016.
[21] F. Yang, J. Cheng, and T. A. Tsiftsis, "Free-space optical communication with nonzero boresight pointing errors," IEEE Trans. Commun., vol. 62, no. 2, pp. 713-725, Feb. 2014.
[22] I. S. Ansari, M.-S. Alouini, and J. Cheng, "Ergodic capacity analysis of free-space optical links with nonzero boresight pointing errors," IEEE Trans. Wireless Commun., vol. 14, no. 8, pp. 4248-4264, Aug. 2015.
[23] N. Sharma, A. Bansal, and P. Garg, "Spectrally efficient TWR-aided free-space optical communication over turbulent channel with generalized pointing error," Trans. Emerg. Telecommun. Technol., vol. 28, no. 4, p. e3052, Apr. 2017.
[24] H. AlQuwaiee, H.-C. Yang, and M.-S. Alouini, "On the asymptotic capacity of dual-aperture FSO systems with generalized pointing error model," IEEE Trans. Wireless Commun., vol. 15, no. 9, pp. 6502-6512, Sep. 2016.
[25] S. Hranilovic, Wireless Optical Communication Systems. Cham, Switzerland: Springer, 2005.
[26] N. Ou, T. Farahmand, A. Kuo, S. Tabatabaei, and A. Ivanov, "Jitter models for the design and test of gbps-speed serial interconnects," IEEE Design Test Comput., vol. 21, no. 4, pp. 302-313, Jul. 2004.
[27] H. G. Olanrewaju and W. O. Popoola, "Effect of synchronization error on optical spatial modulation," IEEE Trans. Commun., vol. 65, no. 12, pp. 5362-5374, Dec. 2017.
[28] L. Yang, K. Panta, and J. Armstrong, "Impact of timing jitter and I/Q imbalance in OFDM systems," IEEE Commun. Lett., vol. 17, no. 2, pp. 253-256, Feb. 2013.
[29] S. Miki et al., "Timing jitter characterization of the SFQ coincidence circuit by optically time-controlled signals from SSPDs," IEEE Trans. Appl. Superconduct., vol. 29, no. 5, Aug. 2019, Art. no. 1302104.
[30] M. T. Dabiri, S. M. S. Sadough, and M. A. Khalighi, "Blind signal detection under synchronization errors for FSO links with high mobility," IEEE Trans. Commun., vol. 67, no. 10, pp. 7006-7015, Oct. 2019.
[31] Y. Li, T. Geng, S. Ma, S. Gao, and H. Gao, "Timing jitter's influence on the symbol error rate performance of the L-ary pulse position modulation free-space optical link in atmospheric turbulent channels with pointing errors," Opt. Eng., vol. 56, no. 3, Mar. 2017, Art. no. 036116.
[32] Y. Li, S. Guo, T. Geng, S. Ma, S. Gao, and H. Gao, "Evaluation on the capacity and outage performance of the free space optical system impaired by timing jitters over an aggregate channel," Opt. Eng., vol. 56, no. 7, Jul. 2017, Art. no. 076108.
[33] A. Ghazisaeidi, "Noise analysis of zero-forcing nonlinear equalizers for coherent WDM systems," J. Lightw. Technol., vol. 37, no. 6, pp. 1552-1559, Mar. 15, 2019.
[34] C. Yang, F. Yang, and Z. Wang, "Iterative minimum mean square error equalization for optical fiber communication systems," IEEE Photon. Technol. Lett., vol. 19, no. 19, pp. 1571-1573, Oct. 1, 2007.
[35] I. Djordjevic, L. Minkov, and H. Batshon, "Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPCcoded turbo equalization," IEEE J. Sel. Areas Commun., vol. 26, no. 6, pp. 73-83, Aug. 2008.
[36] N. Alic, M. Skold, P. A. Andrekson, M. Karlsson, and S. Radic, "Benefits of joint statistics in MLSD-equalized transmission," IEEE Photon. Technol. Lett., vol. 21, no. 8, pp. 495-497, Apr. 5, 2009.
[37] M. Aharonovich and S. Arnon, "Performance improvement of optical wireless communication through fog with a decision feedback equalizer," J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 22, no. 8, pp. 1646-1654, 2005.
[38] A. Bhardwaj, N. Sauer, L. Buhl, W. Yang, L. Zhang, and D. T. Neilson, "An InP-based optical equalizer monolithically integrated with a semiconductor optical amplifier," IEEE Photon. Technol. Lett., vol. 19, no. 19, pp. 1514-1516, Oct. 1, 2007.
[39] N. Kim and H. Park, "Low-complexity iterative equalisation and decoding for wireless optical communications," IET Commun., vol. 2, no. 1, pp. 61-65, 2008.
[40] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," IEEE Trans. Inf. Theory, vol. 18, no. 3, pp. 363-378, May 1972.
[41] N. D. Chatzidiamantis, G. K. Karagiannidis, and M. Uysal, "Generalized maximum-likelihood sequence detection for photon-counting free space optical systems," IEEE Trans. Commun., vol. 58, no. 12, pp. 3381-3385, Dec. 2010.
[42] C. Gong and Z. Xu, "Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference," IEEE Trans. Wireless Commun., vol. 14, no. 10, pp. 5326-5337, Oct. 2015.
[43] SerialXpress Advanced Jitter Generation for AWG. Accessed: 2020. [Online]. Available: https://www.tek.com/product-softwareseries/serialxpress-arbitrary-waveform-generator-software
[44] A. Khatoon, W. G. Cowley, and N. Letzepis, "Channel measurement and estimation for free space optical communications," in Proc. Austral. Commun. Theory Workshop, Melbourne, VIC, Australia, Jan. 2011, pp. 112-117.


Yatian Li (Member, IEEE) received the B.S. and M.S. degrees (Hons.) in communications engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2013 and 2015, respectively, and the Ph.D. degree in mechanical and electrical engineering with the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2019. He is currently a Research Assistant with the Laboratory of Laser Communication, Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun, China. His research interests include free-space optics and wireless communications.


Tianwen Geng received the B.S. and M.S. degrees in electronic circuits and systems from Jilin University (JLU), Changchun, China, in 2004 and 2007, respectively. He is currently an Associate Researcher with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun. His current research interests include free-space optics and fiber communication.

Ruotong Tian received the B.S. degree in electric information science and engineering from Jilin University (JLU), Changchun, China, in 2017. She is currently pursuing the Ph.D. degree in circuits and systems with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun, and also with the University of Chinese Academy of Sciences (UCAS), Beijing, China. Her research interests include coherent detection and digital signal processing.


Shijie Gao received the B.S. degree in mechanical and electrical engineering from HUST, Harbin, China, in 2002, and the M.S. and Ph.D. degrees in electronic circuits and systems from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2006 and 2015, respectively. He is currently an Associate Researcher with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOMP), Chinese Academy of Sciences (CAS), Changchun, China. His current research interests include free-space optics and laser technology.


[^0]:    ${ }^{1}$ There are two main reasons for dividing $\xi_{k}$ into the sign $\Xi\left(\xi_{k}\right)$ and absolute value $\varsigma_{k}$. First, each state in our TEVE algorithm contains the signs' information. In addition, the absolute value can simplify the later derivation process.

[^1]:    ${ }^{2}$ The other situation is $\xi_{k+1}<-1$, which can be discussed by the similar means.
    ${ }^{3}$ Here we ignore the situations of $\xi_{k+2}=0$ and $\xi_{k+2}=\xi_{k+1}-1$. Because these strict equalities are nearly impossible for Gaussian variables $\xi_{k+1}$ and $\xi_{k+2}$.

[^2]:    ${ }^{4}$ The maximum output power of the MZM is 100 mW .
    ${ }^{5}$ The ACPS is plays a role of simulating the power fluctuation by dynamically adjusting the attenuation value. It can be either placed in the transmitter or in the receiver, as long as it is put in front of the photodiode. The ACPS was designed by our laboratory.

